UNIVERSITÉ PARIS DESCARTES

ÉCOLE DOCTORALE

« Sciences humaines et sociales : cultures, individus, sociétés »

THÈSE

pour l'obtention du

DIPLÔME DE DOCTORAT

en sciences de l'éducation, didactique des mathématiques

présentée et soutenue publiquement le 17 novembre 2011 par

Françoise GAYDIER

SIMULATION INFORMATIQUE D'EXPÉRIENCE ALÉATOIRE ET ACQUISITION DE NOTIONS DE PROBABILITÉ AU LYCÉE

Thèse dirigée par Mme Sylvette MAURY

Jury :

M. Eric BRUILLARD, rapporteur

M. Yves CHEVALLARD

Mme Sylvette MAURY

M. Jean-Claude RÉGNIER, rapporteur

Remerciements

Je remercie Madame Sylvette Maury de m'avoir un jour suggéré d'engager un travail sur les simulations d'expériences aléatoires, et d'avoir bien voulu diriger cette thèse.
Sa grande ouverture d'esprit lui a permis tout à la fois de me guider dans ce domaine qui lui est cher, l'enseignement des probabilités, et de m'accompagner dans le domaine de l'algorithmique-programmation.
Elle m'a souvent aidée à distinguer mes activités de recherche de mes activités d'enseignante, et m'a soutenue et encouragée à persévérer dans les moments difficiles.

Je remercie Messieurs Eric Bruillard et Jean-Claude Régnier, qui ont accepté d'être les rapporteurs de ce travail de thèse, et qui me font l'honneur de faire partie du jury.
Je remercie également Monsieur Yves Chevallard de l'honneur qu'il me fait en acceptant de faire partie de ce jury.

Je remercie mes collègues et amies Jacqueline Lhoyer et Véronique Slovacek-Chauveau de leur complicité active : sans elles la partie expérimentale de cette thèse n'aurait pas été possible.

Je remercie toutes les personnes qui m'ont aidée à des titres divers à mener à bien ce travail, en particulier Isabelle Perrin, et les membres du laboratoire EDA.

Je remercie mes amis et mes proches de leurs encouragements au long de ces années, et tout particulièrement Anne-Marie, Michèle et Pierre, Jean-Pierre, Frédéric, Pascale, Françoise et Jacques, Chantal et Bernard.

Je remercie enfin Rafael pour sa compréhension, et Carlos, pour son soutien, et son infinie patience dans la vie quotidienne.

Résumés et mots clés

Les programmes affirment que simuler une expérience aléatoire, c'est simuler sa loi de probabilité : nous montrons que ce n'est pas une nécessité.
Nous avons entrepris une analyse a priori de la tâche de simulation d'une expérience aléatoire lorsqu'on ne la fait pas dériver de sa loi de probabilité ; cela nous a amenée à préciser les liens qu'entretiennent une expérience aléatoire, ses modélisations probabilistes et ses simulations.
Nous proposons un modèle de ces liens, où intervient une étape de *pré-modèlisation*, commune aux deux tâches de simulation et de modélisation probabiliste, étape pendant laquelle sont choisies les *hypothèses de modélisation*. La simulation peut alors se construire à partir d'un *cahier des charges* qui décrit les différentes actions constituant l'expérience aléatoire et leur enchaînement, pour une *imitation au plus près* cette expérience. La simulation informatique apparaît alors comme une activité essentiellement de type algorithmique.
Nous avons mené une expérimentation auprès de lycéens pour observer quelles techniques ils mettent en œuvre pour simuler une expérience aléatoire, et dans quelle mesure ils utilisent le modèle probabiliste ou des simulations pour résoudre un problème de prise de décision dans une situation où intervient le hasard.
Une fois choisies les hypothèses de modélisation, l'*imitation au plus près* n'utilise pas la théorie des probabilités. Certains problèmes résolus par une exploitation statistique des simulations peuvent donc permettre d'introduire des notions de la théorie des probabilités telles que : *risque, intervalle et niveau de confiance, adéquation d'un modèle probabiliste aux données expérimentales.*

Mots clés : expérience aléatoire, probabilité, modèle probabiliste, hypothèses de modélisation, simulation, loi des grands nombres, algorithme.

Computer simulation of random experiments and acquisition of probability concepts in high school

The mathematics curriculum claims that simulating a random experiment amounts to the simulation of the underlying probability distribution. We show here that this is not a necessity.

We performed an a priori analysis of the task of simulating a random experiment without deriving it from its probability distribution. This led us to clarify the connections between a random experiment, its probabilistic modeling and its simulations.

We propose a model for these links with a pre-modeling step, which is common to the simulation and probabilistic modeling tasks and during which the modeling assumptions are chosen. A simulation which imitates closely the random experiment can then be constructed from specifications that describe its list of actions and how they are linked to each other. Computer simulation appears as a core activity of algorithmic type.

We conducted an experiment with high school students to observe which techniques they implement to simulate a random experiment and how they use the probabilistic model or simulations to solve a problem of decision making in situations involving randomness.

Once the hypotheses of the model are chosen, the imitation of the random experiment does not make use of probability theory. Some problems solved by a statistical study of simulations may therefore help to introduce the concepts of probability theory such as risk, confidence interval, level of confidence, relevance of a probabilistic model to experimental data.

Keywords : random experiment, probability, probabilistic model, modeling assumptions, simulation, law of large numbers, algorithm.

SOMMAIRE

INTRODUCTION

1. L'origine de cette thèse

Jusqu'aux années 2000 l'enseignement des probabilités en lycée, même lorsque les programmes préconisaient plutôt une approche fréquentiste, était envisagé par les enseignants presque exclusivement dans une approche laplacienne, avec, selon les programmes en vigueur, la progression suivante :

• dans un premier temps est menée l'étude d'expériences aléatoires simples, modélisées par une loi équirépartie sur un ensemble fini d'issues, l'uniformité de la loi étant suggérée par la symétrie des issues (lancer d'une pièce, d'un dé, tirage au hasard d'une urne). La probabilité d'un événement est alors le rapport du nombre d'issues favorables au nombre d'issues possibles.

• Ensuite des expériences aléatoires plus complexes sont étudiées : calculs de probabilités avec un ensemble d'issues non équiréparties, la loi de probabilité est déduite de données de l'énoncé (« on a constaté que la probabilité qu'une pièce issue de la machine A soit défectueuse est de 0,3 »), ou bien est obtenue par « transport » par une variable aléatoire définie sur un espace probabilisé muni d'une loi équirépartie (par exemple la somme des points obtenus en lançant deux dés équilibrés).

• Enfin, lorsque la notion de probabilité conditionnelle figure au programme, il est possible d'étudier des expériences aléatoires *composées*, c'est à dire constituée d'expériences aléatoires *élémentaires* telles que décrites précédemment, simultanées ou successives, en s'appuyant sur la formule des probabilités totales. La mise en œuvre de la formule des probabilités totales s'écrit *linéairement*, c'est à dire sans utiliser les arbres de probabilité.

Le plus souvent, l'expérience globale proposée à l'étude était essentiellement virtuelle : elle était décrite et rarement réalisée, même lorsque les programmes suggéraient d'appuyer le cours de probabilité sur une étude statistique menée « pour de vrai », pour faire apparaître une relative stabilisation de la fréquence de chaque issue lorsqu'on augmente la taille de l'échantillon des données observées.

La loi des grands nombres et sa preuve ont pu figurer dans les programmes (programme dit des « maths modernes » en Terminale C dans les années 1975), mais en conclusion, on pourrait même dire en apothéose, du chapitre « Probabilités ».

En dehors de cette période, jusqu'aux années 2000, l'énoncé de la loi des grands nombres ne figurait pas explicitement dans les programmes : le lien entre les expériences aléatoires réelles que les programmes recommandaient de faire avec les élèves, les fréquences des issues alors observées, et la probabilité de ces issues, restait à l'initiative de l'enseignant, qui se trouvait souvent bien démuni, tant du fait de sa formation initiale en probabilités insuffisante, que par la difficulté supposée à mettre effectivement en place dans la classe une expérience aléatoire à répéter.

Alors qu'existait, particulièrement au sein des IREM, un fort courant pour une approche fréquentiste de la notion de probabilité, approche suggérée par les programmes, la pratique de l'enseignant dans sa classe pouvait donc être très variable.

Les programmes des années 2000 (applicables à la rentrée 2000 pour la classe de seconde, la rentrée 2001 en Première, et la rentrée 2002 en Terminale) préconisent explicitement une introduction « fréquentiste » de la probabilité d'un événement aléatoire, approche s'appuyant sur « un énoncé vulgarisé » de la loi des grands nombres, que nous citerons au chapitre 2.

Ces mêmes programmes introduisent des activités de simulation d'expériences aléatoires simples, et ce dès la classe de seconde, dans le chapitre de statistique.
Les simulations suggérées sont basées sur « des tables de hasard », ou sur des nombres aléatoires générés par un dispositif informatique (calculatrice, tableur).
Ces programmes de 2000, comme leurs documents d'accompagnement, sont discrets quant à l'auteur de la simulation (le professeur ou l'élève ?).

Lors des consultations des enseignants qui ont accompagné la conception, puis la publication de ces programmes, des enseignants de notre établissement se sont étonnés de la présence de simulations d'expériences aléatoires dans les programmes, s'interrogeant sur leur utilité pour faire comprendre la notion de probabilité et / ou sur leur intérêt dans des situations aléatoires pour lesquelles le modèle probabiliste est facile à établir.
De fait, la notion de fluctuation d'échantillonnage venait d'être introduite dans le programme de statistique de seconde, mais n'était probablement pas encore une notion « en acte ». Par

ailleurs, le statut des nombres pseudo-aléatoires n'était pas encore bien clair.
Enfin les enseignants, encore peu nombreux, qui utilisaient dans leur classe l'outil informatique, utilisaient plutôt des logiciels de construction géométrique que des tableurs [1] ou des logiciels possédant des fonctions aléatoires faciles d'accès. Les logiciels de construction géométrique pouvaient contenir des fonctions aléatoires, mais elle n'étaient ni d'un accès, ni d'un maniement faciles (Cabri-géomètre, Géoplanw ; Géogébra, logiciel plus récent et logiciel libre donc susceptible d'évoluer « facilement », a une fonction aléatoire d'accès non direct peu expliquée et offrant peu de possibilités).

Le lien ne s'est pas fait spontanément, ni pour nous-même, ni pour nos collègues, entre cette apparition de simulations dans les programmes, et les possibilités qu'offrait l'outil informatique.
Cependant des stages ont été organisés pour les enseignants, dans lesquels des formateurs montraient, entre autres, des feuilles de calcul simulant la répétition N fois d'une expérience aléatoire simple, et des présentations de graphiques liés à ces feuilles de calcul : en abscisse la taille n de l'échantillon ($1 \leq n \leq N$) et en ordonnée la fréquence d'une issue lors des n premières épreuves ; ces graphiques pouvaient être utilisés pour montrer la stabilisation de la fréquence d'un événement lorsque l'expérience est répétée un grand nombre de fois, mais aussi, en relançant les calculs, la fluctuation d'échantillonnage.
Pendant cette période, des efforts ont été faits dans les équipement informatiques des lycées, et la nécessité de matériels dédié aux mathématiques s'est imposée : des vidéoprojecteurs ont été installés, permettant de projeter à la classe des images de telles feuilles de calcul.

En 2006 a été annoncée la création d'une épreuve pratique en mathématiques [2] au baccalauréat, épreuve centrée sur l'utilisation de l'outil informatique dans un contexte mathématique. Cette épreuve devait être expérimentale à la session 2007, puis généralisée à la session 2008.
En fait cette épreuve est restée expérimentale en 2008 et 2009, a été suspendue en 2010 et 2011 et sa mise en œuvre effective a été repoussée à la session 2013.

---

[1] en dehors des enseignants en Première L « confrontés » à un enseignement dit « mathématiques et informatique » pour lequel une part de l'activité est centrée sur le tableur, mais pas dans des situations aléatoires

[2] De telles épreuves sont déjà en place en sciences physiques et en SVT, comptant pour un cinquième de la note dans ces disciplines au baccalauréat.

Cependant l'Inspection de mathématiques a fait preuve de beaucoup de volontarisme pour imposer cette épreuve, même expérimentale, dans les lycées, et nous pensons que cette épreuve annoncée a pu changer certaines pratiques dans l'enseignement des probabilités.
En effet, les quelques sujets expérimentaux de cette épreuve pratique qui relèvent du chapitre probabilité, ont exigé des candidats qu'ils construisent une feuille de calcul simulant l'expérience aléatoire à étudier.

Le cadre de l'enseignement des probabilités en classe de première et terminale (S, ES, et L pour les élèves ayant choisi une spécialité maths) est donc, jusqu'à la session 2012 [3], du baccalauréat le suivant :

• approche fréquentiste de la notion de probabilité, avec explicitement un « énoncé vulgarisé » de la loi des grands nombres,
• pour des expériences composées, utilisation d'un arbre de probabilité pour décrire les issues et pour calculer les probabilités,
• simulations d'expériences aléatoires avec des outils informatiques permettant de répéter un grand nombre de fois la simulation,
• la simulation pouvant être l'œuvre de l'élève.

A l'origine de notre thèse il y a le questionnement (évoqué plus haut) qui s'est exprimé lors de la consultation des enseignants sur les nouveaux programmes : qu'est-ce que des simulations d'expériences aléatoires peuvent bien apporter à l'enseignement des probabilités ?

Nous avions l'impression que l'état de la réflexion en la matière se résumait à ceci : l'approche fréquentiste doit être privilégiée ; et puisque l'ordinateur permet de répéter un très grand nombre de fois l'expérience aléatoire (ou plutôt sa simulation), il y a là la possibilité d'illustrer concrètement la loi des grands nombres, en montrant aux élèves la stabilisation de la fréquence vers une fréquence théorique, la probabilité [4].
Il nous semblait alors que le scénario décrit ci-dessus relevait plus de l'intime conviction des prescripteurs que d'une analyse fondée sur des résultats théoriques et (ou) sur des pratiques

[3] Au delà, les programmes sont sensiblement modifiés, nous y reviendrons dans le corps de notre travail.

[4] Rappelons que dans les années 2000, les documents d'accompagnement des programmes insistaient sur le caractère expérimental des mathématiques. Le champ des probabilités, avec les simulations, donne une certaine force à cette conception.

observées dans des situations d'enseignement des probabilités.
Cette impression était plutôt confortée par ce que nous observions dans nos classes dans notre pratique professionnelle : si la projection de feuilles de calcul simulant un sondage d'opinions semblait assez bien reçue comme illustration de la fluctuation d'échantillonnage, nous n'avions pas l'impression que les graphiques illustrant la stabilisation de la fréquence de réalisation d'un événement lorsque la taille de l'échantillon augmente emportaient l'adhésion des élèves. Tout au plus obtenions-nous une attention polie.
Nous pensions que cet échec venait peut-être de la situation de passivité dans laquelle se trouvaient les élèves devant des simulations informatiques présentées par l'enseignant, et que les simulations ne pourraient, peut-être, présenter d'intérêt dans l'enseignement des probabilités que si les élèves en étaient eux-mêmes les auteurs.

Nous avons été convaincue de la nécessité d'essayer de regarder de plus près les enjeux des simulations dans l'enseignement des probabilités par la lecture d'un article de J.P. Delahaye paru en 2005 dans « Pour la Science » (numéro 336, p 90-94).
Dans cet article, que nous présenterons brièvement au chapitre 4, Delahaye explique que, dans des situations aléatoires paradoxales, quelques simulations permettent d'emporter l'adhésion au calcul théorique des chances.
Or d'une part, à aucun moment l'auteur ne fait allusion à la fluctuation d'échantillonnage, qui semble pourtant être un sérieux obstacle pour emporter la conviction de néophytes sur la base de *quelques* simulations.
D'autre part la simulation proposée dans l'article pour convaincre d'un résultat paradoxal obtenu par le calcul dans un problème appelé « l'énigme des Sophies » semble discutable.

Nous avions donc, quand nous avons commencé cette thèse trois objectifs :
- préciser ce que peut être une simulation d'une expérience aléatoire, au moins au niveau de l'enseignement secondaire ;
- analyser la tâche qui consiste, pour un élève, à créer une simulation d'une expérience aléatoire (toujours au niveau du lycée);
- essayer d'observer dans quelle mesure l'élève intègre, dans sa démarche pour résoudre un problème où intervient le hasard, la possibilité de faire des simulations.

Notons qu'il s'est ajouté un nouvel enjeu : un enseignement de l'informatique en tant que discipline va se mettre en place en terminale S à la rentrée scolaire 2012, s'appuyant sur une initiation à l'algorithmique-programmation assurée en classes de seconde et première S par les enseignants de mathématiques. Il apparaît que les simulations informatiques d'expériences aléatoires pourraient avoir une place de choix dans les exercices, problèmes ou projets proposés dans ce cadre aux élèves. Parallélement, la formation des élèves à l'algorithmique-programmation permettra peut-être de proposer aux élèves de construire des simulations d'expériences aléatoires plus complexes.

Notre recherche à l'origine était résolument un travail dans le champ de l'enseignement des probabilités et de la statistique, et il l'est resté. Nous n'avons cependant pas pu faire l'économie d'aborder l'aspect *informatique* des simulations informatiques étudiées, et d'évoquer quelques problèmes didactiques dans ce champ disciplinaire, mais notre préoccupation principale demeure centrée sur l'enseignement des probabilités.

2. La définition des simulations

Notre interrogation sur ce qu'on entend par « simulation » d'une expérience aléatoire, s'est trouvée renforcée par le constat que les programmes, les documents d'accompagnement officiels de ces programmes, et divers travaux proposent avec beaucoup de constance une définition (*simuler une expérience aléatoire, c'est simuler sa loi de probabilité*), mais que cette définition est contredite d'une part par la formulation même des sujets de l'épreuve pratique évoquée au paragraphe précédent, et d'autre part par notre pratique.

Nous avons donc été amenée à tenter d'étudier de très près la manière dont nous-même procédons lorsque nous faisons une simulation informatique d'une expérience aléatoire, étude introspective en quelque sorte, et aussi à analyser les travaux de recherche en didactique où sont en jeu des simulations d'expériences aléatoires, travaux peu nombreux jusqu'à présent, et qui, peut-être, fondent cette définition qui semble si consensuelle.

La loi de probabilité d'une expérience aléatoire ayant un nombre fini k d'issues (de probabilités respectives n_1/N, n_2/N, …, n_K/N avec $n_1 + n_2 + \ldots + n_k = N$) est très facile à programmer : cette loi de probabilité est aussi celle de l'expérience aléatoire qui consiste à tirer une boule d'une urne composée de N boules de k couleurs différentes, n_1 boules étant d'une couleur, n_2 d'une autre couleur, etc., tirage où l'on s'intéresse à la couleur de la boule tirée.

Il nous est apparu que les simulations évoquées dans les travaux de recherche sont toujours organisées autour de tirages d'urnes, et dans des *logiciels fermés*, ce qui semble tout à la fois légitimer la définition des simulations évoquée plus haut, qui est datée (la fin du XX^{ème} siècle), et rendre aujourd'hui nécessaire de la réviser, dans un sens moins restrictif.
Précisons que nous entendons par *logiciel fermé* un programme informatique qui n'est pas conçu par l'élève : l'intervention de l'élève se limite aux choix de certains paramètres, lors de l'utilisation du logiciel, afin de constituer un *modèle d'urne* [5], paramètres qui sont le nombre d'urnes, leur composition, et le type de tirage : avec ou sans remise.

La situation en ce début du XXI^{ème} siècle est toute autre. Le taux d'équipement, en ordinateurs personnels dans les familles, et en terminaux divers dans les lycées, a considérablement augmenté, et surtout, et, c'est peut-être lié, les programmes de mathématiques des années 2010 introduisent un enseignement d'algorithmique-programmation en classes de seconde et première , prolongée par une spécialité « Informatique et Science du Numérique » (ISN) en classe de terminale S à la rentrée 2012.
De ce fait, les élèves acquièrent la faculté **de programmer eux-mêmes les simulations** des expériences aléatoires qu'ils ont à étudier.
L'enjeu n'est plus du tout le même : on passe de la nécessité de produire, sous forme de logiciels fermés, un ersatz plausible d'expériences aléatoires dont la loi de probabilité est connue, à **une tâche dévolue à l'élève**, avec des objectifs pas très clairement définis (on a pu penser, lors de l'expérimentation d'une épreuve pratique de mathématiques en terminale S, que les simulations sur tableur demandées dans les exercices de probabilité proposés aux élèves, avaient plus pour finalité de légitimer cette épreuve « pratique » en salle informatique, que de participer à la résolution d'un exercice de probabilité).

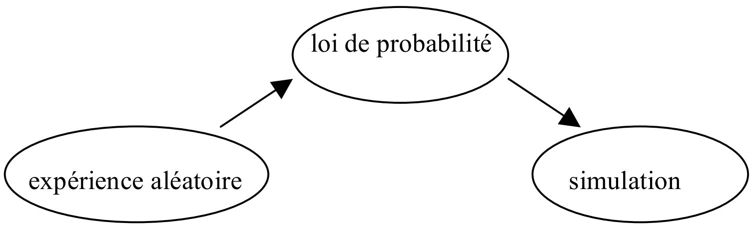
Il n'y a pas – à notre connaissance – de travaux analysant cette tâche de simulation d'une expérience aléatoire dévolue à l'élève.
Nous avons voulu faire une analyse a priori de cette tâche, en nous appuyant d'abord sur notre

---

[5] notion définie de la même manière par Zaki (1990) et Coutinho (2001), mais pas utilisée de la même manière dans leurs expérimentations. Déterminer un *modèle d'urne* consiste à définir la composition d'une ou plusieurs urnes, et le type de tirages (avec ou sans remise) à effectuer. Il s'agit bien sûr de déterminer un *modèle d'urne* « adapté » au problème à résoudre.

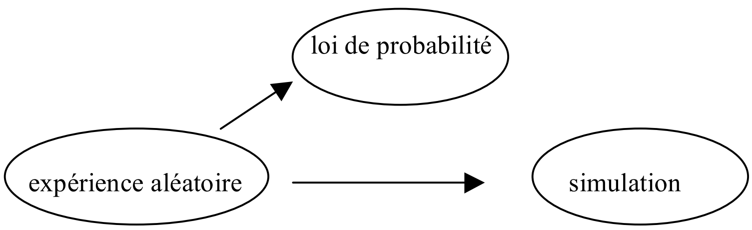
propre pratique et quelques *observations spontanées* [6] dans nos classes.
Notre pratique, comme nous l'avons indiqué, lorsque nous construisons une simulation d'une expérience aléatoire, qu'elle soit simple ou plus complexe, étant rarement la simulation de sa loi de probabilité, nous avons été amenée à essayer de préciser les liens qu'entretiennent les trois sommets de ce triangle « expérience aléatoire / loi de probabilité / simulation » : doivent-ils être représentés par le schéma 1 (i.e. : la définition consensuelle évoquée plus haut) ou par le schéma 2 ?

<u>schéma 1</u>



La *simulation* est alors la simulation de la loi de probabilité, c'est à dire concrètement la simulation d'un tirage sans remise d'une urne dont la composition est définie par la loi de probabilité (cf. plus haut).

<u>schéma 2</u>



Pour ce schéma, la nature du lien représenté par la flèche qui va de l'expérience aléatoire à sa simulation reste à définir, l'existence – ou non – simultanément d'un lien entre la simulation et la loi de probabilité est également à éclaircir.

Les programmes informatiques (dont l'exécution est une simulation de l'expérience aléatoire initiale) que l'on peut trouver dans les documents accompagnant les « nouveaux programmes » (programmes de 2010) confortent le schéma 2.
A partir donc de notre pratique de modélisations probabilistes et de simulations

---

[6] Nous appellerons *observations spontanées* des observations que nous avons pu faire à plusieurs reprises dans des classes, pendant un cours, mais ce ne sont pas des expérimentations.

d'expériences aléatoires, de nos observations spontanées dans nos classes, et de quelques expérimentations limitées dans nos classes sur des points particuliers, nous avons établi un modèle des liens entre les trois sommets du triangle, et nous formulons l'hypothèse que les liens sont en fait plus complexes que ne le font ressortir les schémas 1 ou 2.
Nous émettons également l'hypothèse que les élèves, en situation d'avoir à accomplir une tâche de simulation d'une expérience aléatoire, procéderaient plutôt selon notre modèle que selon le schéma 1 préconisé par les programmes.

3. L'expérimentation

Nos observations spontanées dans nos classes nous ont amenée à formuler une troisième hypothèse, à savoir qu'il est possible d'envisager, **au lycée**, la détermination du modèle probabiliste d'une expérience aléatoire, ou sa simulation, non comme un exercice constituant une fin en soi, mais comme des outils de résolution de ce que nous aurions envie d'appeler, par opposition aux exercices académiques que sont les exercices proposés au baccalauréat, de « vrais problèmes » de probabilité : par exemple un problème de prise de décision dans une situation complexe où intervient le hasard.
Nous considérons en effet, à la suite de nombreux auteurs, en particulier Brousseau et Maury, qu'un problème de probabilité doit présenter un enjeu, un jugement ou une prise de décision.

Nous avons confronté les hypothèses que nous venons d'évoquer aux observations que nous avons pu faire lors d'une expérimentation ; pour mener cette expérimentation, nous avons imaginé un problème, que nous avons soumis à 157 élèves de cinq classes de première et terminale S.

A l'issue de cette expérimentation, certaines observations nous ont amenée à approfondir la réflexion sur l'apport des simulations dans l'enseignement des probabilités : nous avons utilisé l'approche praxéologique de Chevallard pour analyser les deux tâches de modélisation probabiliste et de simulation d'une expérience aléatoire (au lycée), et cette grille de lecture nous a permis d'étudier jusqu'à quel point la simulation d'une expérience aléatoire peut être mise en œuvre en dehors de la connaissance de la théorie des probabilités.

Cette étude nous a amenée à envisager les simulations informatiques, qui ont comme spécificité de pouvoir être répétée un très grand nombre de fois (les résultats observés pouvant

être traités aisément statistiquement), pas seulement comme des substituts d'expériences aléatoires permettant d'illustrer la théorie des probabilités enseignée.
Nous avons alors tenté de préciser ce que des simulations informatiques d'expériences aléatoires, utilisées comme outil dans la résolution d'un problème, peuvent apporter, non plus seulement en terme de connaissances sur l'environnement du problème à résoudre, mais en terme de compréhension des problématiques de la théorie des probabilités : incertitude, contrôle de cette incertitude, prise de risque, intervalles et niveaux de confiance, adéquation du modèle probabiliste choisi pour une expérience aléatoire, etc.

4. Organisation de la thèse

Notre thèse est composée de quatre parties.

Dans la première partie, nous présentons les concepts et les outils que nous avons utilisés pour mener notre travail, qui s'inscrit comme un travail de recherche en didactique des mathématiques : outre les travaux en didactique des probabilités, nous nous appuyons particulièrement sur la théorie anthropologique du didactique de Chevallard et les praxéologies telles qu'il les a définies.
Nous présentons également, dans cette première partie, divers travaux qui nous ont permis de faire, en quelque sorte, un état des lieux du domaine des mathématiques qui nous intéresse ici : les probabilités.
Cet état des lieux concerne les conceptions scientifiques du hasard, les fondements des probabilités, ainsi que les questions concernant l'enseignement des probabilités.
Nous avons fait un détour par les travaux d'Augustin Cournot sur l'applicabilité du calcul des probabilités, en nous appuyant sur divers auteurs : les *principes de Cournot* nous ont paru pouvoir éclairer l'énoncé fluctuant de la *loi des grands nombres* dans les programmes au fil des ans.
Enfin nous présentons, dans cette première partie, les programmes des années 2000, qui seront en vigueur jusqu'à la session 2012 du baccalauréat pour les terminales, et les orientations des nouveaux « nouveaux programmes » que nous appellerons « programmes 2010 ».

Dans la deuxième partie, nous proposons un premier modèle de la tâche « création d'une simulation d'une expérience aléatoire », au niveau de l'enseignement secondaire toujours, et

un modèle des liens entre l'expérience aléatoire, le modèle probabiliste qu'on en propose et la simulation de cette expérience aléatoire.
Ce modèle n'est pas tout à fait en accord avec la définition des simulations donnée dans les programmes : il l'élargit plus qu'il ne la contredit ; ce modèle paraît correctement décrire les activités de simulations qu'il semble raisonnable d'aborder au lycée.

Dans la troisième partie, nous présentons une expérimentation menée dans trois classes de première S et deux classes de Terminale S pendant l'année scolaire 2008-2009 et l'année scolaire 2009-2010.
Les élèves ont eu à effectuer un travail (papier et crayon) de simulation et de modélisation probabiliste d'une situation aléatoire liée à un problème, et à imaginer une voie pour essayer de résoudre ce problème.
Nous avons pu analyser une partie des réponses au travers d'une grille précisant les procédures mises en jeu par les élèves.
Une partie des réponses était plus qualitative. Nous avons tenté d'analyser et de classifier ces réponses, et de les rapprocher du troisième objectif que nous avions fixé à notre thèse : observer dans quelle mesure l'élève intègre, dans sa démarche pour résoudre un problème où intervient une expérience aléatoire, la possibilité de faire des simulations ou de s'appuyer sur une modélisation probabiliste de cette expérience.

Notre quatrième partie a été très largement inspirée par les réponses proposées par les élèves pour résoudre le problème qui leur était posé dans cette expérimentation.
Nous approfondissons d'abord la résolution du problème lui-même, résolution à l'aide de simulations informatiques, et nous analysons les résultats obtenus avec cette résolution, à la lumière de la théorie des probabilités.
Cette analyse nous amenant à discerner, dans cette résolution, ce qui relève des probabilités de ce qui relève de l'algorithmique et de la statistique descriptive, nous avons entrepris de définir des praxéologies pour les deux tâches *modélisation d'une expérience aléatoire* et *simulation d'une expérience aléatoire*, et de faire ressortir le type de connaissances que l'une et l'autre des tâches permettent d'obtenir sur l'expérience aléatoire initiale à étudier.
Nous montrons que des résultats obtenus avec les simulations informatiques d'expériences aléatoires – sans avoir à mettre en œuvre des connaissances de la théorie des probabilités – permettent d'aborder des problématiques spécifiques aux probabilités, et partant peut-être de contribuer à l'acquisition de ce que certains appellent l'*esprit probabiliste*.

Après avoir essayé d'analyser ce que de nombreux auteurs entendent lorsqu'ils évoquent une démarche spécifique en probabilités, nous proposons quelques *types de problèmes* dont la résolution est possible à l'aide de simulations informatiques et d'une étude statistique des résultats que ces simulations produisent.
La simulation n'y est alors plus conçue comme un prétexte à étudier une expérience aléatoire, la simulation (en lieu et place de l'expérience simulée, qu'on s'empresse d'oublier), pour illustrer le cours classique de probabilités.
La simulation devient une étape dans une démarche scientifique pour résoudre un problème dans une situation où intervient le hasard.

La question de la qualité de la simulation, c'est à dire de ce qui permet d'affirmer que l'exécution d'un programme informatique est bien une simulation d'une expérience aléatoire donnée est difficile ; nous avons essayé de la résoudre, en sortant du « cercle vicieux didactique » évoqué par certains auteurs.

La simulation informatique produisant des connaissances sur l'expérience d'origine qu'elle simule, peut constituer un modèle de cette expérience, au même titre que le modèle classique reposant sur la loi de probabilité. La place de ce modèle informatique pourrait être mieux prise en compte dans les programmes et dans la formation des enseignants, et nous espérons que notre travail ouvre quelques voies dans ce sens.

PREMIERE PARTIE

L'enseignement des probabilités.

Note : de nouveaux programmes de mathématiques sont apparus pendant que nous écrivions cette thèse. La « position » des simulations d'expériences aléatoires a sensiblement été modifiée dans ces nouveaux programmes. Nous avons essayé de prendre en compte ces changements, même si nous ne pouvons pas encore en prendre toute la mesure, l'ensemble des nouveaux programmes et des documents les accompagnant n'étant pas encore publié.
Cependant dans la mesure du possible, nous signalerons l'apport de ces nouveaux programmes. Ayant donc à évoquer deux programmes, il nous a paru commode de nommer l'un et l'autre : « les programmes 2000 » et « les programmes 2010 ».

CHAPITRE 1

repères théoriques

Pour une présentation historique du calcul des probabilités et de son enseignement, nous renvoyons par exemple à Courtebras ((2006) et (2008)) [7] et aux thèses de Coutinho (2001), Nabbout (2006) et Dhieb (2009) ; on peut trouver les grandes lignes de l'histoire de la statistique dans la thèse d'Oriol (2007), et les grandes lignes de l'histoire de l'enseignement de la statistique dans la thèse de Wozniak (2005).

Nous ne présenterons pas un aperçu historique des probabilités et de leur enseignement, un tel aperçu ne pourrait que recouper très fortement ceux des travaux citées.
Cependant une telle mise en perspective de la statistique et des probabilités permet de cerner et comprendre les conceptions en jeu, les ruptures qui ont pu se produire, et la classification du hasard que nous évoquerons dans ce chapitre.

Nous allons essayer de préciser le sens que l'on peut donner à certains termes ou concepts que nous serons amenée à utiliser dans nos analyses. Pour certains d'entre eux, nous signalerons simplement que nous nous inscrivons dans un contexte de recherche en didactique, et que nous utilisons ces termes avec le sens communément accepté dans certains cadres théoriques bien identifiés que nous citerons.

Nous serons amenée à parler d'algorithmique dans la deuxième partie de notre thèse. Nous donnerons alors quelques éléments théoriques, et nous contenterons dans ce premier chapitre de donner quelques définitions sommaires.

1. Connaissances, conceptions, rapport au savoir

Il existe plusieurs façons de parler des connaissances d'un individu dans le champ de la didactique des mathématiques, et ces « façons » correspondent à des cadres théoriques différents. On parle aussi de *conceptions*, d'*intuitions*, de *rapport au savoir* d'un sujet.

---

[7] travaux critiqués par Chevallard et Wozniak : note critique publiée dans le numéro 156 de la Revue Française de Pédagogie.

1.1. Intuitions, conceptions spontanées

Fischbein (1975) s'est intéressé aux probabilités intuitives [8], en précisant le lien entre les *intuitions*, qu'il associe à l'action immédiate, et les connaissances : il considère qu'elles sont de même nature, que les intuitions sont des « pré-notions », des « précurseurs » de la connaissance scientifique.

Il constate au travers d'expérimentations que les performances observées pour certaines des situations proposées semblent se dégrader avec l'âge : l'*intuition* serait correcte chez les plus petits, puis serait dégradée chez les plus grands, qui auraient une tendance à l'interprétation déterministe des phénomènes, tendance à laquelle conditionnerait l'enseignement.

Fischbein pense qu'on peut modifier les intuitions probabilistes des enfants en développant plus tôt « le mode de pensée spécifique des probabilités » afin de contrebalancer cette tendance. Nous présenterons de façon plus détaillée ce point de vue de Fischbein dans la quatrième partie.

Les probabilités demeurent l'une des parties des programmes que les enseignants de lycée ont tendance à repousser à la fin de l'année scolaire ou à ne pas traiter : objet « d'allègements (officiels ou non) » écrivait déjà Maury dans sa thèse en 1986. Il n'est plus question d'allègements officiels, mais l'analyse que faisait alors Maury paraît pouvoir être reconduite :

> « *(...) la situation de la statistique et des probabilités est analogue à celle de la mécanique classique (...). Cette analogie pourrait provenir de ce que la statistique et les probabilités, tout comme la mécanique (et la physique plus généralement) entretiennent des liens étroits avec la réalité. Elèves et étudiants aborderaient alors l'apprentissage scolaire de ces disciplines avec un corps de connaissances spontanées important (...) s'inscrivant parfois en contradiction avec les connaissances scientifiques. Ainsi en probabilités – plus, peut-être, que dans d'autres secteurs des mathématiques où "l'expérience première" jouerait un rôle moins important – la construction systématique des concepts, au sein du système scolaire, ne peut ignorer la* <u>*dotation "spontanée"*</u> *ou "intuitive" des élèves.* » Maury(1986)

Nous parlerons de *conceptions spontanées* d'une notion par un élève pour évoquer une conception de cette notion que s'est construit cet élève avant le processus d'enseignement de

[8] *probabilités objectives*, *probabilités subjectives* et *probabilités intuitives* : cf. paragraphe 5.2. de ce chapitre.

cette notion. De telles conceptions spontanées peuvent être *correctes* ou *erronées*.

Une conception peut être correcte ou erronée, et chez un même individu peuvent coexister, pour un même objet, des conceptions correctes et des conceptions erronées.
Nous allons tenter de décrire diverses acceptions de ce mot « conception »

1.2. Concepts et conceptions

Balacheff (1995 b) écrit :

> « *Le mot "conception" est utilisé depuis de nombreuses années dans les recherches sur l'apprentissage et l'enseignement des mathématiques, mais il fonctionne "sans qu'au départ les auteurs semblent éprouver le besoin d'en donner une définition didactique" (Artigue 1991, p.266).* Conception *fonctionne ainsi comme un concept " paradidactique " : il est présent, de fait, dans les travaux de recherche comme un outil mais sans accéder au rang d'objet d'étude, pour lui-même. Dans une acception dominante, comme le montre Artigue, le terme de conception renvoie à un "objet local" (...)* »

Maury et Caillot (2003) écrivent :

> *« D'une certaine façon, on peut* [donc] *dire que la notion de conception, qui prend en considération à la fois le fonctionnement du sujet mais également celui du savoir, assure un lien entre la psychologie et la didactique. »*

Dans ce même ouvrage (« Rapport au savoir et didactiques »), Maury et Caillot rappellent qu'au début des années quatre-vingt, les chercheurs en didactique « *considéraient l'élève comme un pur sujet épistémique isolé* ».

L'étude des conceptions des élèves consistait souvent en une comparaison de leurs savoir-faire avant et après l'enseignement.
Cependant il y a des conceptions erronées qui peuvent provenir d'un choix pédagogique. Maury et Caillot (2003) citent la conception "un décimal = un entier" résultant de l'introduction des décimaux à partir des mesures de grandeur :

> « *En définitive, dans l'exemple cité, il apparaît que les conceptions ne sont pas structurées par la seule étude des conduites du sujet. Il est clair qu'elles sont aussi*

largement déterminées par l'analyse des conditions d'élaboration du savoir. »

Vergnaud (1991) a donné une définition *pragmatique* de la notion de *concept* :

« Un concept est un triplet de trois ensembles, C= (S, I, ζ)

- S, l'ensemble le des situations qui donnent sens au concept (la référence) ;

- I, l'ensemble des invariants sur lesquels repose l'opérationnalité des schèmes (le signifié) ;

- ζ, l'ensemble des formes langagières et non langagières qui permettent de représenter symboliquement le concept, ses propriétés, les situations et les procédés de traitement (signifiant). »

Balacheff (1995 a) s'appuie sur cette définition pour présenter une *caractérisation formelle* du concept de *conception* ; une conception est alors une caractérisation d'un état du système sujet/milieu :

« *Nous appelons conception C, un quadruplet (P, R, L, ∑) dans lequel :*
- P est un ensemble de problèmes sur lequel C est opératoire ;
- R est un ensemble d'opérateurs ;
- L est un système de représentation, il permet d'exprimer les éléments de P et de R ;
- ∑ est une structure de contrôle, elle assure la non contradiction de C.
En particulier, un problème p de P est résolu si il existe r de R et s de ∑ tel que s(r(p))=vrai.
Un opérateur transforme un problème en un nouveau problème, une condition minimale pour qu'un problème appartienne à P est qu'il existe une suite de transformations par des éléments de R qui conduise à un problème résolu au sens de ∑. »

Pour Balacheff, les structures de contrôle sont implicites dans la définition de Vergnaud :

« *Elles sont, par exemple, impliquées de fait par le recourt au concept de théorème-en-acte ou celui d'inférence qui n'ont de sens qu'associés à la reconnaissance chez le sujet de procédures permettant de s'assurer de la légitimité et de la correction de la mise en oeuvre des conduites correspondantes*. »

Il précise (Balacheff (1995 b) :

> « *Cette caractérisation n'est pas plus du côté du sujet, que du côté du milieu avec lequel il interagit. Elle permet au contraire une caractérisation d'un état du système sujet/milieu : le système de représentation doit permettre d'exprimer et de mettre en oeuvre les opérateurs tant du point de vue du sujet émetteur-actif que du milieu récepteur-réactif, les structures de contrôle doivent tout autant modéliser les moyens du sujet-acteur que la capacité du milieu-réacteur à assurer des rétroactions discernables.* »

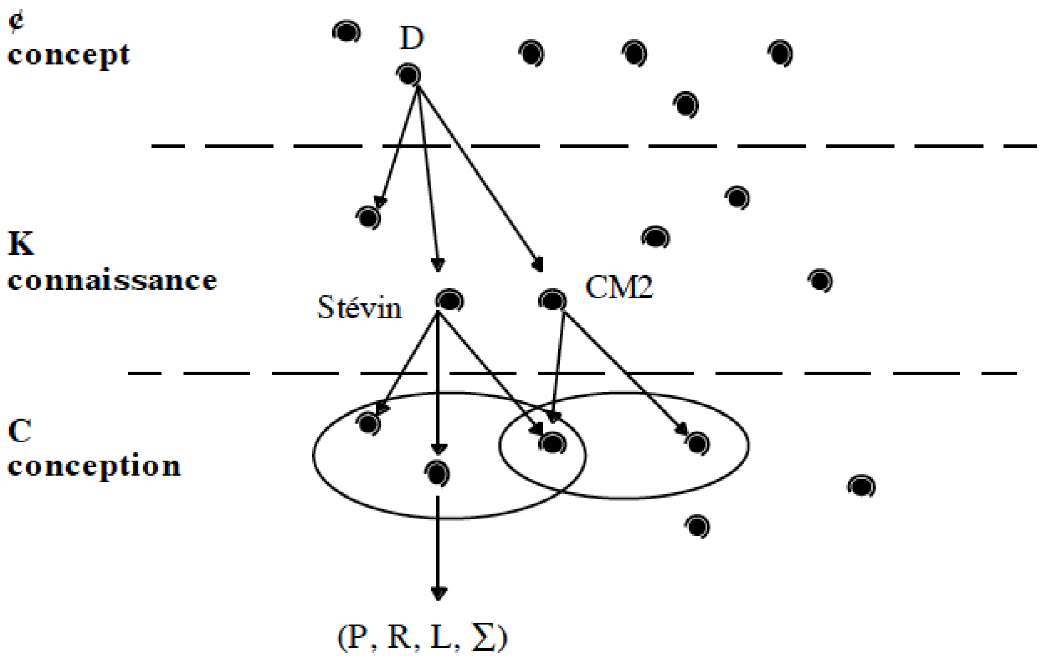
Balacheff « remonte » de sa définition formelle de *conception* vers la définition de *concept* : il définit entre deux conceptions $C=(P, R, L, \sum)$ et $C'=(P', R', L', \sum')$ les relations « C est plus générale que C' » et « C est fausse au sens de C' » à partir d'une *fonction de représentation* f : $L' \rightarrow L$.

Il définit une conception particulière C_μ plus générale que toutes les autres (son existence est « *un postulat dont il importe peu qu'il ait directement une instance dans la réalisation de notre modèle* »).

Si C et C' sont deux conceptions distinctes, et si f et f' sont les fonctions de représentation où $f : L \rightarrow L_\mu$ et $f' : L' \rightarrow L_\mu$, C et C' ont le même µ-objet (ou un même objet au sens de C_μ) si pour tout p appartenant à P, il existe p' de P' tel que $f(p)=f'(p')$, et réciproquement.

Une *connaissance* est alors un ensemble de conceptions ayant le même µ-objet, et un *concept* l'ensemble des connaissances ayant le même µ-objet.

Balacheff (1995 b) donne le schéma suivant :



Ce shéma résume la structure que nous proposons. Le concept de "nombre décimal" (D) est déterminé par son extension, les connaissances D-selon-Stevin, D-en-CM2. Chaque connaissance est déterminée par un ensemble de conceptions qui sont chacunes caractérisées par un quadruplet (P, R, L, Σ).

Il remarque :

> *« Une conception est une instanciation de la connaissance d'un sujet par une situation (elle caractérise le couple sujet/milieu en situation), ou encore une instance d'un concept par un couple sujet/situation. »*

On a pu observer diverses évolutions dans les modèles proposés par les chercheurs en didactique : pour permettre la prise en compte du système sujet/milieu avec Balacheff, ou la prise en compte du contexte (ou des contextes) dans lequel ont lieu les apprentissages, et les effets de ce contexte.

1.3. Rapport au savoir

Un savoir « vit » dans plusieurs institutions. Il faut entendre par *institution* aussi bien la famille, l'école, la classe, le milieu professionnel, la religion, le langage etc.

Une institution a un rapport à un certain objet de savoir : on parlera de *rapport institutionnel au savoir*.

Un individu a une relation avec cet objet de savoir : on parlera de *rapport personnel au savoir*.

Un individu appartient à diverses institutions, et c'est au travers de ces diverses institutions qu'il entre en rapport avec le savoir.
Le *rapport personnel au savoir* d'un individu est donc lié au rapport à ce savoir qu'entretiennent les diverses institutions auxquelles appartient cet individu.

Cette approche est qualifiée d'anthropologique et a été développée par Chevallard (2003) dans sa théorie anthropologique du didactique (T.A.D.)
Dans ce cadre de la T.A.D., la notion de *rapport au savoir* se substitue à la notion de *connaissances* ou de *conceptions*.

Les théories ne s'opposent pas : elles ne sont pas organisées autour des mêmes questionnements.
Maury et Caillot (2003) établissent d'ailleurs un parallèle entre les notions de *conception* et de *rapport personnel au savoir*, rapport personnel au savoir qui peut avoir une « composante publique » et une « composante privée » :

> « *Ainsi, en quelque sorte cet élève affirme à l'expérimentatrice que ce qu'il dévoile dans l'institution "expérimentation" n'est pas la même chose que ce qu'il donnerait à voir dans l'institution "cours de mathématiques".*
> *Cette réponse, ou les réponses de ce type, peut être interprétée dans le cadre théorique des conceptions. C'est d'ailleurs ce que nous faisons dans le travail cité (Maury, 1987), l'idée étant qu'un individu dispose généralement de plusieurs conceptions relatives à une même notion et que les conceptions sont mobilisées de manière sélective en fonction d'éléments contextuels* »

Chevallard (2003) formalise la notion de « connaissance » à partir des notions fondamentales d'*objet*, de *rapport personnel d'un individu à un objet*, de *personne* et d'*institution* :
- le *rapport personnel* de l'individu à l'objet o est « *le système noté R(x,o) de toutes les interactions que x peut avoir avec l'objet o* ».
En particulier, o existe pour x si R(x,o) n'est pas vide ce qu'on note $R(x,o) \neq \varnothing$. Chevallard dit alors que *x connaît o*, et le rapport R(x,o) indique *la manière dont x connaît o.*
- Une personne est le couple formé par un individu x et le système de ses rapports personnels <u>à un moment donné</u> : le système des rapports R(x,o) change ; la personne évolue, l'individu x est invariant.
L'*univers cognitif* de x est l'ensemble $U(x) = \{(o, R(x,o))/R(x,o) \neq \varnothing\}$

- L'individu x appartient à diverses institutions. Pour un objet o, une institution I et une position p dans I, Chevallard définit *le rapport institutionnel à o en position p* : $R_I(p,o)$.
Par exemple I peut être la classe de première S, ou la classe de première S5, p peut être le professeur de la classe ou un élève de la classe ; et o peut être une équation du deuxième degré.
L'individu x est un *bon sujet de I en position p* si R(x,o) est <u>conforme</u> à $R_I(p,o)$.

Chevallard parle alors de « la relativité institutionnelle de la connaissance » : R(x,o) peut être conforme à $R_I(o,p)$ et ne pas être conforme à $R_{I'}(o,p')$.

> « *La relativité institutionnelle de la connaissance est marquée à la fois par l'existence d'une diversité pratiquement illimitée de façons de " connaître " un objet o et par l'inexistence d'un "bon rapport" universel, reconnu tel en toute institution* ».

Pour un objet o et une institution I, et un individu x occupant la position p dans I, si R(x,o) est jugé non conforme à $R_I(p,o)$, on décide <u>dans I</u>, que x « ne connaît pas » o, ou que x « connaît mal » o.

2. Réalité, modèles et théories

Le terme « modèle » renvoie à divers cadres théoriques ; nous en ferons usage avec divers sens, que nous serons amenée à préciser. Lorsque ce sera possible, nous lui adjoindrons un épithète ou l'insérerons dans une locution, après avoir défini la locution complète ; nous parlerons ainsi de *modèle probabiliste*, *modèle informatique*.

La phrase « Jacqueline est mon modèle » ne se comprendra pas de la même façon, selon qu'elle est énoncée par Picasso (qui est en train de peindre Jacqueline) ou d'une admiratrice d'une certaine Jacqueline (que l'admiratrice s'efforce d'imiter).
Dans un cas, le modèle est la réalité, c'est Jacqueline : le peintre va en abstraire des éléments, afin de représenter cette réalité ; dans l'autre, le modèle est une abstraction de la réalité, une sélection de certaines propriétés de Jacqueline, que l'admiratrice tente d'incarner aussi : le style vestimentaire, la façon de chanter, le mode de vie, etc.

On retrouve peut-être cette ambiguïté avec l'expérience aléatoire qui consiste à lancer un dé cubique. On dit que la loi équirépartie sur l'ensemble {1 ; 2 ; 3 ; 4 ; 5 ; 6} est un modèle de cette expérience (parmi peut-être d'autres modèles possibles).

Mais ne pourrait-on dire aussi que le *lancer d'un dé équilibré* est un modèle de la loi équirépartie sur {1 ; 2 ; … ; 6}, comme on parle de *modèle réduit* d'une automobile par exemple, pour dire que l'objet ressemble à cette automobile ?
En effet, parler du lancer d'un dé équilibré renvoie immédiatement au modèle de la loi équirépartie sur {1 ; 2 ; 3 ; 4 ; 5 ; 6} :

> *« Convenons de dire que le terme « dé équilibré », parle du modèle associé au lancer de dé et dit que ce modèle est la loi équirépartie. Cela peut être démotivant pour l'élève de « supposer » le dé équilibré (en maths, on ne peut rien faire normalement…), surtout si l'enseignant à du mal à répondre à la question : comment sait-on s'il est équilibré ?* ***Le dé est un outil très ancien de simulation de la loi équirépartie*** [9] *: pourquoi mettre en doute d'entrée sa raison d'être ? »* Schwarz (2011)

On peut dire la même chose pour le *tirage au hasard* d'un jeton parmi 6 jetons numérotés de 1 à 6, mais aussi, dans une certaine mesure, pour l'appel de la fonction « Alea.entre.bornes(1 ; 6) » du tableur, expériences qui *simulent* aussi la loi équirépartie. Dans ce contexte où l'on rapproche ces trois expériences aléatoires, on bute sur la définition qu'il convient de mettre derrière le mot *simulation* : c'est une *simulation* de la loi équirépartie parce que « ça lui ressemble », comme le *modèle réduit* ressemble à l'automobile qu'il copie ?

En fait, l'évocation du lancer d'un *dé équilibré*, ou d'un *tirage au hasard* renvoie souvent plus au modèle (la loi équirépartie), qu'à la réalisation <u>effective</u> de cette expérience aléatoire. Un dé est-il jamais véritablement équilibré ? A la limite ne pourrait-on pas plutôt considérer que l'évocation de chacune de ces trois expériences est <u>synonyme</u> de l'évocation de la loi équirépartie sur {1 ; 2 ; … ; 6} ?
Cela vaut aussi pour l'appel de « Alea.entre.bornes(1 ; 6) » si son évocation renvoie au tirage d'un entier entre 1 et 6 selon la loi équirépartie, plutôt qu'à l'implémentation du calcul de nombres pseudo-aléatoires dans un langage de programmation.

---

[9] souligné par nous.

A propos du lancer d'un dé (équilibré) ou du tirage (au hasard) d'une urne, Henry (2001) parle de *modèles pseudo-concrets*. Dans ce cadre théorique, l'évocation du tirage d'une urne, renvoie à la modélisation de l'expérience aléatoire à étudier : une modélisation à l'aide d'un *modèle d'urne*, préparant à la notion de probabilité (cf. la thèse de Coutinho (2001), chapitre 5, paragraphe 3) ou permettant des *estimations* (cf. la thèse de Zaki (1990), chapitre 5, paragraphe 1).

Cependant, si le lancer d'un dé ou le tirage d'une urne peut être associé par l'élève à une expérience aléatoire « concrète », cette autre expérience aléatoire qui consiste en l'appel de « Alea.entre.bornes(1 ; 6) » a des chances de rester plutôt « abstraite » pour un élève de lycée, en dehors de l'évocation du modèle de la loi équirépartie sur {1 ; 2 ; … ; 6}, quand bien même il aurait reçu quelques explications sur les nombres pseudo-aléatoires.

2.1. modèles et théorie dans la *théorie des modèles*

On peut également faire un rapprochement avec la *théorie des modèles* de la logique, et considérer que ces trois expériences aléatoires sont trois modèles, trois instanciations, trois réalisations d'une même théorie : la loi équirépartie sur {1 ; 2 ; … ; 6}.

En effet, en logique mathématique, plus précisément dans la *théorie des modèles* , un modèle de la théorie T est un ensemble qui satisfait les axiomes de cette théorie T, et par conséquent les théorèmes de cette théorie.
Ainsi, l'ensemble des nombres réels, l'ensemble des nombres complexes, sont des modèles de la théorie des corps ; l'ensemble des nombres réels est un modèle de la théorie des corps totalement ordonnés complets, mais n'est pas un modèle de la théorie des corps algébriquement clos (ce qu'est l'ensemble des nombres complexes).
Dans ce contexte, le modèle est **une** « incarnation », une réalisation, une instanciation d'une théorie, éventuellement parmi d'autres incarnations : une théorie, comme nous venons de le voir, peut avoir plusieurs modèles.

Par analogie avec la *théorie des modèles*, le lancer d'un dé cubique équilibré, le tirage au hasard d'un jeton parmi six jetons numérotés de 1 à 6, l'appel d'une fonction aléatoire pourraient se comprendre comme des *modèles* de la loi équirépartie {1 ; 2 ; 3 ; 4 ; 5 ; 6}.
Les analogies peuvent être dangereuses, et il n'y a pas ici de *théorie* au sens de la *théorie des*

modèles ; nous verrons (chapitre 7, paragraphe 5) que la présentation que font Schwartz (2011) et d'autres auteurs du *lancer du dé équilibré* comme <u>*simulation*</u> *de la loi équirépartie* soulève la question de la définition que l'on donne de *simulation d'une expérience aléatoire*.

<u>2.2. modèles et réalité</u>

Nous renvoyons à la thèse de Dhieb (2009), p. 24 à 28, pour des développements sur les modèles en sciences humaines. Il présente une définition générale de la notion de modèle en citant Walliser (1997) :

> « *Dans sa définition la plus large, la notion de modèle recouvre toute représentation d'un système réel, qu'elle soit mentale ou physique, exprimée sous forme verbale, graphique ou mathématique* »

La définition est très large, puisqu'elle inclut aussi bien le modèle réduit que construit le modéliste, que le modèle de la désintégration nucléaire présenté conjointement en classe de terminale scientifique par les enseignants de mathématiques et de sciences physiques conformément aux programmes des années 2000.

Cette définition ne pose pas la finalité du modèle, à la différence de celle de Julo (1995), cité par Dhieb :

> *« un modèle possède deux caractéristiques principales : il est une simplification d'un système donné et il permet une action sur ce système. »*

Chevallard (1992) précise les notions de *théorie* et *modèle*, en rejetant « *l'illusion représentationnaliste* », qui présenterait les théories et les modèles comme des *représentations* ou des *images* du système dont on s'occupe.
Chevalllard considère que la « *métaphore culturelle de l'image constitue un véritable obstacle épistémologique* » pour définir la notion de modèle. Il poursuit, dans un paragraphe intitulé *Les métaphores comme outil de pensée* :

> « *Contrairement à ce que beaucoup pensent, j'affirme ici que toute activité scientifique (y compris en mathématiques) se constitue (en son langage) et se décrit (en son métalangage) par l'usage de* ***métaphores****. (…)*
> *Il n'est donc a priori nullement illégitime de penser théories et modèles en termes d'image et de représentation. Mais le grand problème, ici, tient dans le choix des "bonnes" métaphores, des métaphores véritablement fécondes, et dans leur*

contrôle. (...)

*Mais revenons au thème de la théorisation. Afin de penser l'activité de théorisation (ou de modélisation), j'introduis maintenant **une autre métaphore**, aussi légitime en son principe que celle de la représentation, mais, je crois, beaucoup plus libératrice.*

*Je dirai donc qu'une théorie ou un modèle d'un système donné est une **machine** dont la mise en fonctionnement permet **de produire des connaissances** relatives au système modélisé (ou théorisé). (Je prends ici le mot connaissance dans son acception courante, en usant donc d'un métalangage accessible à partir de la culture courante, au contraire du langage théorique que je présenterai ensuite).* »

Chevallard précise, comme Julo, qu'un modèle a une finalité : l'action (résoudre un problème pour Julo, *produire des connaissances* chez Chevallard), et il aborde le lien entre le modèle (ou la théorie) et la réalité (le système modélisé), qui, dit Chevallard, ne s'analyse pas en terme de ressemblance.

> *« On voit d'abord que la métaphore de la machine à produire des connaissances permet de se débarrasser de l'obsession trompeuse de la "ressemblance" entre théorie et réalité théorisée. (Une machine permettant de produire une certaine pièce n'a aucune raison de ressembler à la pièce à fabriquer, non plus qu'à l'objet dans lequel cette pièce viendra prendre sa place) »* (idem)

<u>2.3. Simulations, modèles probabilistes et système réel de l'expérience aléatoire</u>

Nous définirons au paragraphe 4 de ce chapitre le *modèle classique* d'une expérience aléatoire, que nous appellerons *modèle probabiliste* ; ce modèle probabiliste est la référence explicite des programmes, l'un des objets de savoir dont l'enseignement commence en classes de première et terminale des lycées dans les « programmes 2000 ».

L'objectif que Glayman et Varga (1973) assignent à la simulation d'une expérience aléatoire, qui consisterait à « *remplacer l'étude de l'expérience par celle d'une table de nombres aléatoires qui **imite** le phénomène* », peut être revisité à la lumière des possibilités qu'offrent les *simulations informatiques*.

Remplacer l'expérience aléatoire par une simulation permet d'estimer la probabilité d'un événement par la fréquence d'apparition de cet événement lorsqu'on répète un grand nombre de fois la simulation ; mais les simulations qui s'appuient sur des tables de hasard, par exemple, sont coûteuses en temps, et de ce fait pas toujours efficaces comme le souligne Bordier (1991), p. 24-25.

D'autres auteurs placent la simulation d'une expérience aléatoire entre réalité et modèle, une modélisation théorique étant un préalable à une simulation. Ainsi Dhieb (2010) :

> « *Nous entendons par simulation la reproduction artificielle d'une expérience aléatoire réelle, cette reproduction étant l'exécution informatique d'un objet mathématique qui est supposé modéliser l'expérience en question. Simuler l'expérience du jet d'une pièce de monnaie revient donc à faire fonctionner, sur calculatrice ou ordinateur, l'équiprobabilité considérée comme modèle adéquat pour cette expérience.* »

Nous tenterons de préciser les liens entre un modèle probabiliste d'une expérience aléatoire, une simulation (informatique ou non) de cette expérience, et la réalité (c'est à dire l'expérience aléatoire elle-même). Nous proposerons dans le chapitre 9 de la deuxième partie un modèle décrivant ces liens et nous verrons que plusieurs techniques permettent de construire une *simulation* d'une expérience aléatoire.
Nous aurons alors à préciser les liens entre une *simulation informatique* d'une expérience aléatoire et cette expérience : les simulations informatiques sont-elles des programmes dont l'exécution produit des résultats entretenant seulement des *ressemblances* avec les issues des expériences aléatoires simulées ?

3. Théories didactiques

Nous ne présentons pas en détail certaines théories didactiques que nous avons déjà évoquées, ou que nous serons amenée à évoquer, que ce soit la théorie des situations didactiques de G.Brousseau ou la théorie de l'anthropologie didactique (T.A.D.) de Y. Chevallard.

Nous aurons à décrire et analyser des *tâches*, relevant de *types de tâches* : tâche de modélisation d'une expérience aléatoire, tâche de simulation d'une expérience aléatoire.
Les tâches de modélisation et de simulation d'une expérience aléatoire peuvent être analysées à l'aide de la notion d'*organisation praxéologique* de Chevallard (1999), afin de faire ressortir que les types de tâches auxquelles appartiennent ces tâches dévolues aux élèves, ne

relèvent pas des mêmes technologies, et partant ne posent probablement pas les mêmes problèmes didactiques, ce que notre thèse va tenter de faire ressortir.

Il semble donc ici nécessaire de présenter succinctement cette notion d'organisation praxéologique, qui selon Chevallard, est à la base de la TAD, théorie qui postule que *« **toute** activité humaine régulièrement accomplie peut être subsumée sous un modèle **unique**, que résume ici le mot **praxéologie** ».*

3.1. Tâches et types de tâches

Une *tâche* t relève d'un *type de tâche* T : par exemple « lire le mode d'emploi de cet appareil photo » est une *tâche* t qui relève du *type de tâche* T « lire un mode d'emploi », « calculer $x^2 + 3x - 1$ pour $x = 4$ » est une tâche t qui relève du type de tâche T « calculer l'image d'un nombre par un trinôme ».

Pour une expérience aléatoire donnée notée EA, on peut définir deux tâches : une tâche t_m qui consiste à déterminer un *modèle probabiliste* de EA, et une tâche t_s qui consiste à construire une simulation de EA.

t_m appartient au type de tâche T_m : *modéliser une expérience aléatoire.*

t_s appartient au type de tâche T_s : *simuler une expérience aléatoire.*

Chevallard (1999) souligne que les tâches comme les types de tâches

> *« ne sont pas des donnés de la nature : ce sont des "artefacts ", des "œuvres", des **construits institutionnels**, dont la **re**construction en telle institution, et par exemple **en telle classe**, est un problème à part entière, **qui est l'objet même de la didactique** »*

Une tâche est à faire dans une institution donnée : les praxéologies sont dynamiques ; elles varient d'une institution à l'autre, elles varient dans le temps.

Par exemple, les praxéologies relatives à T_m et T_s ne sont peut-être pas les mêmes pour l'élève que pour le professeur ; elles ne sont peut-être pas les mêmes en l'an 2000 qu'en l'an 2010 du fait des changements de programme.

La praxéologie relative à un type de tâche T indique (en principe remarque Chevallard) comment réaliser les tâches t appartenant à T : *une manière d'accomplir* t ; *«(...) à une telle **manière de faire**, τ, on donne ici le nom de **technique** ».*

3.2. Techniques

Une praxéologie relative au type de tâches T (dans une institution donnée rappelons-le) contient un *bloc* [T/ τ] que Chevallard nomme *bloc pratico-technique,*

> *« qu'on identifiera génériquement à ce qu'on nomme couramment **un savoir-faire** : un certain type de tâches, T, et une certaine manière,* τ, *d'accomplir les tâches de ce type ».*

En général, une technique τ ne permet pas de réaliser toutes les tâches t de T : elle a une *portée* qui est une partie P(τ) de T.

Par ailleurs deux techniques τ_1 et τ_2 peuvent permettre de réaliser des tâches de l'intersection de P(τ_1) et P(τ_2) : une technique peut être supérieure à l'autre.

Nous aurons à nous poser cette question particulièrement pour le type de tâches T_m défini plus haut : « simuler une expérience aléatoire » dans l'institution I qu'est la classe, ou plus précisément dans l'institution I qu'est un cours de probabilité dans une classe de première S suivant les programmes des années 2000.

Chevallard remarque que :

> *« Enfin, en une institution I donnée, à propos d'un type de tâches T donné, il existe en général **une seule** technique, ou du moins **un petit nombre** de techniques **institutionnellement reconnues**, à l'exclusion des techniques alternatives possibles – qui peuvent exister effectivement, mais alors en d'**autres** institutions. Une telle exclusion est corrélative, chez les acteurs de I, d'une illusion de "naturalité" des techniques **institutionnelles** dans I – faire ainsi, c'est naturel… –, par contraste avec l'ensemble des techniques alternatives possibles, que les sujets de I ignoreront, ou, s'ils y sont confrontés, qu'ils regarderont spontanément comme **artificielles**, et (donc) "contestables", "inacceptables", etc. A cet égard, on observe assez fréquemment chez les sujets de I, de véritables **passions institutionnelles** pour les techniques naturalisées dans l'institution. »*

3.3. Technologies et théories

A une technique τ pour réaliser des tâches de T est associée une technologie θ, qui est un *discours rationnel* sur τ, qui a pour fonction à la fois de *justifier* τ , en particulier que τ permet bien de réaliser les ou un ensemble de tâches de T, et d'*expliquer* (au sens : *rendre intelligible, éclairer*) pourquoi.

Une troisième fonction d'une technologie correspond plus à l'acception courante de ce mot :

il s'agit de *produire* des techniques. Chevallard [1999] note :

> *« (...) il existe des technologies* **potentielles**, *en attente de techniques, qui ne sont encore technologies d'aucune technique ou de très peu de techniques. »*

Nous aurons à nous poser cette question de l'existence de *technologies potentielles* pour le type de tâche « modéliser une expérience aléatoire ».

La technologie a, à son tour, besoin d'être justifiée à un autre niveau, *de justification-explication-production* dit Chevallard : celui de la *théorie*, Θ, qui est à θ ce que θ était à τ.
En général, les trois niveaux : technique/technologie/théorie suffisent pour rendre compte de l'activité que l'on veut analyser. Et même souvent, les deux premiers niveaux suffisent, la justification théorique étant renvoyée à une autre institution censée, comme dit Chevallard, détenir cette justification.
Cela peut s'exprimer en classe de mathématiques par : « on admet que », ou « on pourrait démontrer que », et en classe de physique par : « on démontre en mathématiques » etc.

Nous nous interrogerons également sur la ou les théories qui sous-tendent les technologies associées au techniques mises en œuvre pour résoudre les tâches des types de tâches T_m et T_s de modélisation probabiliste et de simulation d'une expérience aléatoire.

3.4. Praxéologie, savoir-faire et savoirs

Si l'on considère un type de tâche T pour lequel on connaît une technique τ au moins, une technologie θ de τ, et d'une théorie Θ de θ, $[T/ \tau / \theta / \Theta]$ est une *praxéologie, ponctuelle* au sens qu'elle est relative à la seule tâche T.
Nous avons déjà dit que le bloc $[T / \tau]$ est un *savoir-faire* ; Le bloc $[\theta / \Theta]$ est identifié comme un *savoir*. Citons à nouveau Chevallard :

> *« Par métonymie, on désigne couramment comme étant un savoir la praxéologie* $[T/ \tau / \theta / \Theta]$ ***toute entière**, ou même une partie quelconque de celle-ci. Mais cette manière de faire encourage à* **minorer le savoir-faire**, *notamment dans la production et la diffusion des praxéologies : ainsi qu'on l'a noté, on rencontre souvent des technologies qui "attendent leur premier emploi", ou qui ont "perdu leur emploi" ! »*

Dans une institution donnée I, une même théorie Θ peut justifier plusieurs technologies θ_j ,

chacune justifiant des techniques τ_{ij} correspondant aux types de tâches T_{ij}.
Les praxéologies ponctuelles vont, comme dit Chevallard, s'agréger en *organisations locales* $[T_i / \tau_i / \theta / \Theta]$ centrées sur la technologie θ, puis en *organisation régionales* $[T_{ij} / \tau_{ij} / \theta_j / \Theta]$ formées autour de la théorie Θ, puis en *organisation globale* $[T_{ijk} / \tau_{ijk} / \theta_{jk} / \Theta_k]$ en agrégeant des théories Θ_k de l'institution I.

Nous retiendrons en conclusion de cette présentation de l'organisation praxéologique cette remarque de Chevallard :

> *« Une organisation praxéologique, même ponctuelle, n'est pas en général entièrement conforme aux canons évoqués ci-dessus. (…) La notion de praxéologie apparaît ainsi comme une notion générique dont il convient d'approfondir l'étude – notamment par l'enquête empirique et l'analyse des données d'observations recueillies. »*

4. La théorie des probabilités : modèle probabiliste

Un *modèle probabiliste* d'une expérience aléatoire est une description de l'expérience aléatoire dans le cadre de la théorie des probabilités telle qu'elle s'est mise en place au 20ème siècle avec Kolmogorov :

• un univers U décrivant toutes les issues possibles pour cette expérience [10].
• une tribu $\mathscr{B}$ définie sur U [11], un élément de $\mathscr{B}$ étant un événement.
• une application probabilité p définie sur $\mathscr{B}$ [12].

On associe à l'expérience aléatoire le triplet (U, $\mathscr{B}$, p) : l'espace probabilisé associé à l'expérience aléatoire, le couple (U, p) étant un espace probabilisable.

Passer d'un espace probabilisable à un espace probabilisé signifie faire le choix d'une application probabilité sur $\mathscr{B}$.

---

[10] U est également nommé « univers des possibles ». Le mot « issue » est parfois remplacé par « résultat possible ».

[11] « tribu » : voir par exemple la définition dans un cours de topologie générale.

[12] C'est à dire une application de $\mathscr{B}$ vers [0 ; 1] satisfaisant :

(i) pour tout A et B <u>disjoints</u> de $\mathscr{B}$, $p(A \cup B) = p(A) + p(B)$

(ii) $p(U) = 1$.

Une fois choisi l'espace probabilisable qui pourra rendre compte de l'expérience à étudier, il y a plusieurs applications probabilités possibles.

Il y a donc, dans la théorie des probabilités, pour une même expérience aléatoire, plusieurs modèles possibles : ils dépendent du *choix* de l'espace probabilisable, puis du *choix* de l'application probabilité qui fera de cet espace un espace probabilisé.

Nous reviendrons dans le chapitre 7 de la deuxième partie sur ces *choix*, qui font partie de l'activité de modélisation de l'expérience aléatoire.

Dans l'enseignement secondaire, en dehors des programmes dits « des maths modernes », la transposition didactique de ce triplet s'est toujours faite dans le cas particulier de la tribu définie par l'ensemble des parties de U : $\mathscr{B} = \mathscr{P}(U)$.

L'événement {w} où w appartient à U est « un événement élémentaire » ou « issue » de l'expérience aléatoire.

Pour déterminer l'application probabilité p, lorsque l'univers U est discret (fini ou non d'ailleurs), il suffit de connaître tous les $p_i = p(\{w_i\})$ pour tous les w_i de U, la probabilité d'un événement A étant alors $p(A) = \sum p_i$, pour tous les w_i constituant A :

> *la probabilité d'un événement est égale à la somme des probabilités de tous les événements élémentaires qui le composent* [13].

L'étude des séries numériques n'étant pas dans les programmes de l'enseignement secondaire, les expériences aléatoires proposées ont un nombre fini d'issues <u>quand l'univers U est un ensemble discret</u>.

Nous verrons dans le chapitre 3 que les « programmes 2000 » tentent une incursion vers les *probabilités continues* , c'est à dire les espaces probabilisés pour lesquels l'univers U est un intervalle réel (lois uniformes sur [a ; b], ou lois exponentielles sur IR^+). Cela suppose bien sûr que le calcul intégral figure dans le programme.

Le lien entre d'une part le cadre théorique des espaces probabilisés, qui peut s'intégrer dans la théorie des ensembles d'un point de vue complètement abstrait, et d'autre part l'étude

---

[13] Cette phrase se trouve, telle quelle ou sous une forme voisine, dans à peu près tous les manuels scolaires de première.

d'expériences aléatoires concrètes, peut [14] se limiter à ce qu'évoquent les noms des objets définis dans ce cadre théorique : « univers des possibles », « issues », « événement », « probabilité », « événement impossible », « événement certain », « événements incompatibles », « événements indépendants », « variable aléatoire », « espérance », « valeur moyenne », tous mots déjà chargés de sens au niveau des connaissances spontanées des élèves.
C'est ce qu'on a pu observer peut-être de façon caricaturale pendant la période des programmes dits « des mathématiques modernes » de la fin des années 1970.

Ni le mot aléatoire, ni le mot hasard ne sont définis dans ce cadre théorique des espaces probabilisés.

L'inégalité de Bienaymé-Tchebycheff [15] se démontre aisément dans ce cadre théorique, à partir de la variance de la loi de probabilité d'une variable aléatoire X [16].
Ce théorème énonce que *la probabilité que X prenne des valeurs éloignées de sa valeur moyenne diminue assez rapidement lorsqu'on s'éloigne de cette valeur moyenne*, puisque lorsqu'on agrandit l'intervalle $]m - t\,\sigma, m + t\,\sigma[$ en augmentant t, la probabilité que X soit en dehors de cet intervalle diminue au moins en raison inverse du carré de t.

La loi des grands nombres se démontre à partir de l'inégalité de Bienaymé Tchebychev et de la loi binomiale, et se démontre sans référence explicite au hasard, en restant dans le cadre de la théorie des ensembles ; la répétition n fois d'une expérience aléatoire où les épreuves (les essais) sont deux à deux indépendants se décline sous l'hypothèse : on travaille dans l'espace probabilisé produit (il s'agit ici d'indépendance « *a priori* », cf. par exemple Maury (1986)).

Nous avons fait ces remarques concernant l'inégalité de Bienaymé Tchebychev et la loi faible des grands nombres pour souligner qu'il est possible de faire une transposition didactique d'une partie de la théorie des probabilités, démonstrations incluses, sans tenter d' « accrocher » cet enseignement à une réalité à étudier, oubliant le caractère « mixte » des

[14] ici, « peut » au sens de « il est possible », ce n'est pas une nécessité…

[15] La probabilité pour que X, de valeur moyenne m et d'écart type σ prenne une valeur en dehors de l'intervalle $]m - t\,\sigma, m + t\,\sigma[$ est inférieure à $\frac{1}{t^2}$, quel que soit t.

[16] Encore faut-il qu'elle existe, ce qui est le cas lorsque U est un ensemble fini.

probabilités, comme de la statistique, en les enfermant dans des « mathématiques épurées » comme dit Wozniak (2005)

La formule des probabilités totales [17] et le théorème de Bayes [18] sont également des théorèmes dont la démonstration est facile (la formule des probabilités totales et sa démonstration figurent en général dans les programmes de terminale S).

5. Algorithme et algorithmique

La notion d'*algorithme* est très ancienne. Nous renvoyons à Nivat (2009) pour une histoire succincte de l'algorithmique.
L'algorithmique entre dans les « nouveaux programmes » de mathématiques (« programmes 2010 ») des lycées.
On lit ainsi pour les programmes de première S (applicables à la rentrée 2011) :

> *« En seconde, les élèves ont conçu et mis en œuvre quelques algorithmes. Cette formation se poursuit tout au long du cycle terminal. Dans le cadre de cette activité algorithmique, les élèves sont entraînés à :*
> *- décrire certains algorithmes en langage naturel ou dans un langage symbolique ;*
> *- en réaliser quelques-uns à l'aide d'un tableur ou d'un programme sur calculatrice ou avec un logiciel adapté ;*
> *- interpréter des algorithmes plus complexes. »* (B.O. spécial n° 9, 30/09/2010)

Nous avons essayé de voir l'évolution du sens du mot *algorithme* et de son dérivé, l'adjectif *algorithmique* au travers de quelques dictionnaires et encyclopédies.
Nous trouvons dans un dictionnaire (Le petit Larousse illustré, 1999) la définition suivante pour le mot algorithme :

> *« Ensemble de règles opératoires dont l'application permet de résoudre un problème énoncé au moyen d'un nombre fini d'opérations. »*

---

[17] Si $B_1, B_2, \ldots, B_n$ forment une partition de l'univers U, alors quel que soit l'événement A,
$p(A) = \sum p(A \cap B_i) = \sum p_{B_i}(A) \times p(B_i)$

[18] Pour faciliter l'énoncé de la formule de Bayes, nous utilisons l'ancienne notation des probabilités conditionnelles : les $B_1, B_2, \ldots, B_n$ formant une partition de l'univers U,
pour tout k entier entre 1 et n, $p(B_k/A) = \dfrac{p(B_k)\ p(A/B_k)}{\sum p(B_i)\ p(A/B_i)}$, la somme se faisant sur l'indice i variant de 1 à n.

Dans le Larousse en 5 volumes (1987) :

« algorithme : n.m., (latin médiéval algorihmus, latinisation du nom d'un mathématicien arabe, avec influence du grec arithmos, nombre). Ensemble de règles opératoires dont l'application permet de résoudre un problème énoncé au moyen d'un nombre fini d'opérations. (un algorithme peut être traduit, grâce à un langage de programmation, en un programme exécutable par un ordinateur). »

Et pour « algorithmique » dans ce même dictionnaire :

« algorithmique : adj.

1. Qui concerne les algorithmes, peut être exprimé par un algorithme.

2. Se dit des langages, tel l'Algol, conçus pour faciliter l'expression d'algorithmes.

3. Musique algorithmique, mode de composition dont les possibles sont calculés au moyen de programmes et de machines électroniques. (Le promoteur principal en est Pierre Barbaud) »

Il ressort particulièrement de ces définitions, maintenant un peu anciennes, qu'il doit y avoir un nombre fini d'opérations à effectuer.

On peut remarquer également que dans la dernière définition, *algorithmique* a le seul statut d'adjectif.

Dans la version 2004 de « Encyclopedia Universalis », on trouve :

« L'objet de l'algorithmique est la conception, l'évaluation et l'optimisation des méthodes de calcul en mathématiques et en informatique. Un algorithme consiste en la spécification d'un schéma de calcul, sous forme d'une suite d'opérations élémentaires obéissant à un enchaînement déterminé. (…)

Divers algorithmes ont été en fait connus dès l'Antiquité dans le domaine de l'arithmétique ou de la géométrie, parmi lesquels, notamment :

- les règles de calcul de longueur d'arcs et de surfaces des civilisations égyptienne et grecque ;

- plusieurs méthodes de résolution d'équations en nombres entiers, à la suite des travaux de Diophante d'Alexandrie au IV^{ème} *siècle ;*

- l'algorithme d'Euclide (env. 300 av. J.-C.) qui permet le calcul du plus grand

commun diviseur de deux nombres ;

- le schéma de calcul du nombre π dû à Archimède.

Ensuite ont été étudiées les méthodes numériques de résolution d'équations algébriques (algorithme de Newton, méthode d'élimination de Gauss), puis d'équations différentielles. Enfin, l'avènement de calculateurs électroniques, à l'issue de la Seconde Guerre mondiale, a entraîné un renouvellement complet de l'algorithmique, appliquée désormais soit à des problèmes de taille bien supérieure à celle des problèmes traités manuellement (d'où la nécessité de méthodes nouvelles), soit surtout à des types de problèmes nouveaux, comme le tri, la recherche rapide d'informations non numériques (base de données) ou l'optimisation combinatoire. »

Le mot *algorithmique* est devenu un adjectif substantivé, ce qui traduit un renversement de point de vue : il y avait des algorithmes, et ce qui avait trait aux algorithmes était « algorithmique ».

Il y a maintenant un domaine, l'algorithmique, qui s'occupe des algorithmes et qui produit des algorithmes.

De quelle nature est ce domaine ? Est-ce une partie des mathématiques ? Est-ce une partie d'une discipline scientifique que serait l'informatique ? L'algorithmique est-elle à leur intersection ?

Et quels sont ses liens avec la logique quand il est question de machines de Türing, de calculabilité, de la thèse de Church ?

Nous ne pouvons répondre à ces questions qu'en faisant des analogies avec d'autres champs de connaissances scientifiques, à leur histoire et à leur genèse : l'algorithmique est elle aussi à la croisée de chemins.

Flajolet et Parizot (2004) ont écrit :

« Même si les algorithmes sont souvent considérés comme étant du ressort exclusif des mathématiques et de l'informatique, leur champ d'application est en réalité beaucoup plus vaste. (...)

Un algorithme, très simplement, c'est une méthode. Une façon systématique de procéder pour faire quelque chose : trier des objets, situer des villes sur une carte, multiplier deux nombres, extraire une racine carrée, chercher un mot dans le dictionnaire (...)

La vertu essentielle des algorithmes est de permettre l'exécution optimisée de procédés répétitifs, essentiellement grâce à la formalisation et à la description des enchaînements logiques à un niveau plus abstrait, et donc plus général. Ils s'étendent ainsi à des domaines de la société toujours plus nombreux et plus inattendus. Cette généralisation a accompagné le développement des langages de programmation depuis les années 60, qui permettent aujourd'hui la manipulation de structures et d'objets ayant des propriétés et des comportements analogues à ceux du monde ordinaire. Leur relation avec les couches profondes du traitement informatique est assurée après coup par le compilateur, de manière transparente pour le programmateur et le concepteur d'algorithmes.
En définitive, le codage numérique des objets manipulés (au niveau informatique) est devenu secondaire pour l'algorithmique. L'essentiel est de percevoir les éléments clés d'un processus de calcul, ou d'un procédé quelconque, et d'imaginer les suites d'opérations logiques les plus astucieuses et les plus efficaces pour le mettre en œuvre de façon automatique et performante. L'algorithme est donc en réalité le squelette abstrait du programme informatique, sa substantifique moelle, indépendante du mode de codage particulier qui permettra sa mise en œuvre effective au sein d'un ordinateur ou d'une machine mécanique.[19] *»*

Si les algorithmes ont existé avant les ordinateurs, les ordinateurs n'existent que pour mettre en œuvre des algorithmes grâce à des *programmes informatiques* qui traduisent l'algorithme en des instructions exécutables par l'ordinateur.
Voilà pourquoi l'algorithmique est associée si étroitement à l'informatique. Mais ce lien entre un algorithme et sa mise en œuvre sur un ordinateur pose des problèmes spécifiques, tels que le temps de calcul ou la taille de la mémoire par exemple qui font que l'algorithmique ne se limite pas à la programmation des algorithmes.

L'enseignement de l'algorithmique s'est d'abord inscrit dans les études d'informatique : dans le monde professionnel, à l'université et dans les grandes écoles.
Dans l'enseignement secondaire a été introduite une *option informatique* (1981 à 1992 et 1995 à 1998) qui comportait un pan important d'algorithmique-programmation.

[19] souligné par nous.

L'algorithmique est à nouveau enseignée dans l'enseignement secondaire (programmes de mathématiques de seconde de 2009), et une spécialité « informatique et science du numérique » (I.S.N.) en terminale scientifique est prévue pour la rentrée 2012, qui comportera également une certaine part d'algorithmique..
Nous serons amenée à évoquer dans notre troisième partie des problèmes didactiques spécifiques concernant l'enseignement de l'algorithmique - programmation.

6. Conceptions scientifiques du hasard et fondements des probabilités

Le contenu d'un enseignement des probabilités dans l'enseignement secondaire est une transposition du « *savoir savant* » dans ce domaine.

Nous allons présenter une classification du hasard telle qu'elle paraît communément acceptée par les scientifiques en ce début de 21ème siècle.
Nous évoquerons ensuite les grandes conceptions vis à vis de la probabilité d'un événement (probabilité objective, probabilité subjective) : nous ne serons plus dans les fondements mathématiques, mais dans les fondements psychologiques, qui déterminent les comportements devant l'aléatoire.
Enfin, nous présenterons la problématique de « l'applicabilité » du calcul des probabilités telle qu'elle a pu être posée par Augustin Cournot.

6.1. Une classification du hasard

Lahanier-Reuter (1999) présente une classification du hasard, fondée en partie sur celle proposée par Benoît Mandelbrot : le hasard du tirage au sort, le hasard bénin, le hasard lent, le hasard sauvage et le hasard formel.

6.1.1. Le hasard du tirage au sort

Il s'agit du hasard lié à l'origine aux jeux ; les problèmes à résoudre étaient au départ essentiellement des problèmes d'équité : en particulier le problème dit des partis résolu par Pascal et Fermat (milieu du 17ème siècle) : comment répartir équitablement les mises posées par les joueurs au début d'une partie d'un jeu de hasard qui se joue en plusieurs tours, lorsque le jeu s'arrête avant la fin de la partie ? Il s'agit de tenir compte du gain que pourrait espérer chaque joueur, au moment où le jeu s'arrête, si la partie continuait.
Dans ce problème ne se pose pas l'étude de ce qui se passe si l'on répète une expérience.

Dans ce type de conception du hasard, ne sont évoqués ni la loi des grands nombres, ni le théorème central limite.

La répétition apparaît cependant indirectement, semble-t-il, dans le problème dit du chevalier de Méré (17ème siècle) : il s'agit encore de hasard lié à des jeux, et de l'observation de ce qui passait pour un paradoxe : en lançant trois dés et en s'intéressant à la somme des points amenés, une certaine somme devait, selon le calcul (erroné) des chances proposé, avoir la même chance qu'une autre, ce que contredisait l'observation des sommes obtenues en répétant l'expérience.
La résolution de ce paradoxe passa par le calcul correct des chances : le calcul resta dans le cadre du hasard du tirage au sort.

6.1.2. Le hasard bénin

Les situations où intervient le hasard bénin ne sont plus seulement des situations liées à des jeux.
Les expériences où intervient le hasard bénin sont reproductibles et la loi des grands nombres, et le théorème central limite sont valides.
Par conséquent, lors de la répétition un certain nombre de fois d'une expérience où intervient ce type de hasard, on observera des fluctuations pour les fréquences des événements qu'on voudra observer, mais pour un très grand nombre d'épreuves de cette expérience, on pourra observer une stabilisation des fréquences.

Le hasard du tirage au sort est un cas particulier de hasard bénin.

Pour ce type de hasard bénin, l'observation d'un petit nombre d'épreuves (échelle microscopique) donne une impression de désordre, une absence de structure, alors que pour un très grand nombre d'épreuves (échelle macroscopique), apparaît une structure, un ordre : « *Ce lien entre phénomènes micro et phénomènes macro est une caractéristique du hasard "bénin"* » dit Lahanier-Reuter (1999), la deuxième caractéristique du hasard "bénin" étant le fait que la dispersion des variations autour de la moyenne tend à suivre une loi normale. De ce fait, la variabilité constatée à l'échelle microscopique est en quelque sorte contenue, puisque à l'échelle macroscopique, les valeurs varient « *d'une manière* *raisonnablement uniforme* » [20] autour de la valeur moyenne.

Lahanier-Reuter poursuit (p.14 et 15) :

> *« Le hasard "bénin" est un hasard raisonnable, puisque les effets extraordinaires sont très peu probables, puisqu'il met en évidence une valeur qui est la plus probable (la moyenne).*
>
> *En articulant le fait que ce hasard révèle le comportement macroscopique, et qu'il le révèle rapidement, par un relevé des valeurs des variables observées, on comprend que ce hasard va être une aide, un outil dans le domaine scientifique. Faire l'hypothèse qu'une variable est une variable aléatoire dans cette optique, n'est pas du tout un obstacle à la description, à la connaissance de cette variable, mais bien au contraire une modélisation efficace. »*

Le hasard de la transposition didactique dans l'enseignement secondaire est implicitement un hasard bénin : les programmes des lycées des dernières décennies évoquent la loi des grands nombres (voir le chapitre 3) (mais pas le théorème central limite). De ce fait, rien dans les programmes n'interdit d'évoquer de façon théorique des variables aléatoires pour lesquelles la valeur moyenne ne serait pas la valeur la plus probable même si les expériences aléatoires concrètes que la prescription propose à l'étude relèvent du hasard bénin.

6.1.3. Le hasard lent

C'est un hasard pour lequel la loi des grands nombres, et le théorème central limite sont encore valides, mais pour lequel l'ordre que révèle l'échelle macroscopique ne se révèle que très lentement, trop lentement.

6.1.4. Le hasard sauvage

Pour ce hasard, ni la loi des grands nombres, ni le théorème limite central ne sont valides. Quelle que soit l'échelle d'observation choisie, on observe le désordre.

Les phénomènes relevant du hasard lent ou du hasard sauvage ne sont pas modélisables avec les outils « classiques » du calcul des probabilité que sont les variables aléatoires et leurs différents moments, dans la mesure où ces moments ne convergent pas ou ne sont pas définis.

[20] J. Gleik, cité par Lahanier-Reuter (1999).

6.1.5. Le hasard formel

Le point de vue n'est plus le même : par rapport à deux suites de 0 et de 1 de même longueur n, codant les résultats obtenus en répétant n fois le lancer d'une pièce, il ne s'agit plus d'expliquer que ces deux suites ont la même probabilité, mais de voir si l'une des suites a l'air plus aléatoire que l'autre (imaginons par exemple que la première ne comporte que des zéros). Le cadre théorique est celui de la théorie de l'information.

Ce cadre permet d'étudier les suites aléatoires selon la quantité de hasard qu'elles contiennent, plus précisément en lien avec la complexité de la suite, la complexité d'une suite pouvant se mesurer à la longueur minimale (longueur en nombre de bits) du programme dont l'exécution engendre cette suite.

Une suite aléatoire, dans ce cadre là, est aléatoire si sa complexité est du même ordre que le nombre de bits qui compose la suite elle-même.

Ce point de vue offre une définition d'une suite aléatoire, mais cette définition n'est pas opératoire : il est possible de démontrer qu'une suite n'est pas aléatoire, mais pas de démontrer qu'une suite est aléatoire.

6.2. A propos des fondements psychologiques des probabilités

Le bref aperçu qui va suivre s'inspire largement de Maury (1986).

Les questions liées à l'aléatoire , en particulier : « sur quoi un individu se fonde-t-il pour gérer son comportement par rapport à l'aléatoire ? » , ont été étudiées par divers auteurs.

On distingue habituellement deux grands courants :

• le courant objectiviste

• le courant subjectiviste.

6.2.1. Le courant objectiviste

Pour ce courant, la probabilité d'un événement est fondée sur deux définitions :

- la probabilité a priori, c'est à dire $\frac{\text{le nombre de cas favorables à cet événement}}{\text{le nombre de cas possibles}}$ lorsqu'on peut faire l'hypothèse que la loi est équirépartie,

- la probabilité empirique, obtenue à partir de la répétition de l'expérience, les épreuves étant indépendantes. On calcule la fréquence de réalisation de l'événement pour un certain nombre d'épreuves, qui est la probabilité empirique.

Ces deux définitions, sous l'hypothèse d'une loi équirépartie, sont mises en cohérence par la loi des grands nombres qui assure que si le nombre n de répétitions de cette expérience est très grand, la probabilité empirique se rapproche de la probabilité a priori.

Si l'on ne peut faire l'hypothèse que les issues de l'expérience aléatoire sont équiprobables, et si l'on ne peut se ramener à cette hypothèse en changeant l'univers des issues (via une variable aléatoire), la probabilité de l'événement est la probabilité empirique pour une répétition « idéalement » une infinité de fois de l'expérience, la probabilité empirique déterminée pour n épreuves se rapprochant de cette probabilité si n est très grand, toujours selon la loi des grands nombres.

La probabilité a priori semble être une conception spontanée précédant l'apprentissage scolaire des probabilités pour des expériences telles que le lancer d'un dé ou d'une pièce, qui sont implicitement supposés équilibrés, ou un tirage « au hasard » d'une urne.

Cette conception de la probabilité d'un événement du courant objectiviste fonde (implicitement) les programmes de l'enseignement secondaire, et est donc la norme à enseigner.

6.2.2. Le courant subjectiviste

Pour ce courant, la probabilité d'un événement n'est que subjective, et se rapporte d'abord à l'état d'esprit du sujet qui évalue cette probabilité, c'est à dire son degré de confiance dans les événements à venir, qui sont incertains.

Il y a deux tendances chez les « subjectivistes » :

• la probabilité est totalement subjective, elle ne peut reposer sur une connaissance objective, et il n'y a pas de régularité pour un même individu non plus qu'entre sujets ;

• la probabilité est intuitive, fondée sur des *intuitions*, qui sont selon Maury (1986) des « *connaissances globales non explicitement justifiées, ressenties par le sujet comme des évidences et s'exprimant dans l'action immédiate* ».

Les valeurs de cette probabilité ne proviennent ni d'un calcul a priori, ni d'une approximation empirique à partir de statistiques conduites sur des événements observés, mais proviennent d'impressions et de jugement subjectifs.

6.3. « L'applicabilité » du calcul des probabilités et le principe de Cournot

Dans la période qui a précédé l'axiomatisation de la théorie des probabilités avec les travaux de Kolmogorov, à la fin du XIX^{e} et au début du XX^{e} siècle, des mathématiciens tels que Jacques Hadamard, Paul Lévy, Augustin Cournot interrogeaient les fondements du calcul des probabilités, peut-être pas tant en termes d'unification des cadres théoriques intégrant les savoir-faire nombreux acquis depuis le 17ème siècle, qu'en termes des liens qu'entretiennent les résultats des calculs de probabilité et la vie réelle, et plus précisément les liens entre ces calculs et l'action, la pratique ; autrement dit ils interrogeaient « *l'applicabilité du calcul des probabilités* » pour reprendre une formulation de Thierry Martin (2007).

Ce qui était recherché là n'était pas du seul ressort individuel (probabilité a priori, probabilité subjective etc, telles que présentées par les psychologues), mais plutôt du ressort institutionnel : une question posée à la communauté scientifique.

Citons Martin (2007) :

> *« Cependant, il ne s'agit pas seulement d'identifier la signification qu'il convient d'attribuer aux applications du calcul des probabilités, mais déjà de savoir si cette application elle-même est possible. Si la question peut aujourd'hui surprendre, tant sont fréquents et variés les recours au calcul des probabilités dans les différentes disciplines scientifiques, elle est posée de manière radicale par d'Alembert (1767) qui, distinguant ce qui vaut mathématiquement et ce qui vaut physiquement, reconnaît la rigueur de la théorie mathématique, mais en conteste l'applicabilité, et elle fait à nouveau l'objet de vives discussions au début du XIX^{e} siècle parmi les mathématiciens et les philosophes.*
>
> *C'est pourquoi, ayant exposé les éléments de la théorie des probabilités, Cournot ajoute aussitôt qu'il convient de se demander si " cette théorie n'est qu'un jeu de l'esprit, une spéculation curieuse, ou si elle a au contraire pour objet des lois très importantes et très générales, qui régissent le monde réel " (1843, §38) »*

Plus loin, Martin explique :

> *« Dans les termes qui sont ceux de Cournot, la question est donc de savoir si la théorie des probabilités peut recevoir une signification objective ou si elle doit s'en tenir à une interprétation subjective. (...) D'une part, en effet, poser que la probabilité reçoit une signification objective, c'est admettre l'existence effective*

de phénomènes aléatoires, et c'est donc s'opposer au postulat d'un déterminisme intégral (…). De l'autre, selon la signification reconnue aux résultats du calcul, la connaissance rationnelle sera considérée soit comme apte à rendre compte adéquatement du réel, soit seulement comme l'instrument permettant d'élaborer la représentation que nous formons à son sujet. »

Cournot ne tranche pas entre une interprétation objective et une interprétation subjective de la probabilité : il assigne un double sens à ce concept, « *qui tantôt se rapporte à une certaine mesure de nos connaissances, et tantôt à une mesure de la possibilité des choses, indépendamment de la connaissance que nous en avons ».* Cournot (1843, p.4 de la préface), cité par Martin.

Cournot tente de préciser ce qu'est le hasard :

> *« Les événements amenés par la combinaison ou la rencontre d'autres événements qui appartiennent à des séries indépendantes* [21] *les unes des autres, sont ce qu'on nomme des événements* ***fortuits*** *ou des résultats du* ***hasard »****.* Cournot (1851), cité par Martin.

Il s'agit en effet de se démarquer du déterminisme laplacien, en affirmant la compatibilité du hasard et du déterminisme :

> *« Les actes des êtres vivants, intelligents et moraux ne s'expliquent nullement, dans l'état de nos connaissances, et nous pouvons hardiment prononcer qu'ils ne s'expliqueront jamais par la mécanique des géomètres. Ils ne tombent donc point, par le côté géométrique ou mécanique, dans le domaine des nombres ; mais ils s'y retrouvent placés, en tant que les notions de combinaison et de chance, de cause et de hasard, sont supérieures dans l'ordre des abstractions à la géométrie et à la mécanique, et s'appliquent aux faits de la nature vivante, à ceux du monde intellectuel et du monde moral, comme aux phénomènes produits par les mouvements de la matière inerte. »*

On nomme volontiers aujourd'hui « <u>principe de Cournot</u> » ce qui soutenait la conception de Cournot du hasard : le concept d'<u>événements physiquement impossibles</u>, tels que l'équilibre

[21] Cournot précisera cette notion d'indépendance. Il définira ailleurs (Cournot [1861]) le hasard comme «*rencontre entre des faits rationnellement indépendants les uns des autres »*.

d'un cône sur sa pointe.
Selon ce principe, si on fixe à l'avance un événement de probabilité nulle ou presque nulle, cet événement ne se produira pas. Par exemple, obtenir une alternance régulière de Piles et de Faces en lançant 100 fois une pièce.

Il y a historiquement la distinction entre *certitude physique* et *certitude morale* : un événement de probabilité très petite est *moralement impossible* et ne se produira pas, et un événement de probabilité très proche de 1 est *moralement certain* et se produira.

Glenn Shafer (2007) donne un panorama historique de ces notions d'événements *moralement certains* ou *moralement impossibles*. Il explique que Bernoulli, ayant démontré la loi des grands nombres, conclut dans *Ars conjectandi* que la très grande probabilité, lorsqu'on répète un grand nombre de fois l'expérience, que la fréquence d'un événement soit proche de sa probabilité, donne une certitude morale et autorise à utiliser la fréquence comme « une indication de sa probabilité ».
Buffon lui met des degrés dans les *certitudes morales* et *physiques*, en lien avec les valeurs des probabilités des événements concernés et construit une échelle de certitude.

Cournot a repris cette idée, mais avec la notion de probabilités infiniment petites : il est possible, mathématiquement, qu'un cône repose en équilibre sur la pointe, mais cela est physiquement impossible car la probabilité de cet événement est infiniment petite.
L'idée est reprise alors, nous dit Shafer, par Bolzman et Henri Poincaré : un événement dont la probabilité est infiniment petite ne se produira pas.
Un enjeu au début du XX^{e} siècle, formulé par exemple dans *die Encyklopädie der mathematischen Wissenschaft* en 1901, était de « *déterminer quel degré de probabilité valait comme certitude pratique afin de pouvoir appliquer la théorie des probabilités* » , Shafer [2007].
Markov en 1912 écrivait :

> *« Plus la probabilité d'un événement est proche de un, plus nous avons de raisons de nous attendre à ce que l'événement se produise et non à attendre son contraire.*
> *Pour les questions pratiques, force nous est de considérer comme certains les événements dont la probabilité est plus ou moins proche de un, et de considérer comme impossibles les événements dont la probabilité est faible.*
> *Par conséquent, l'une des tâches les plus importantes de la théorie de la*

probabilité est d'identifier les événements dont la probabilité est proche de un ou de zéro. » (cité par Shafer)

Emile Borel, au début du 20ème siècle, s'appuyant toujours sur le principe de Cournot, dit du résultat d'un calcul de probabilité qu'il est objectif si sa probabilité est tellement grande qu'elle équivaut à une certitude, et il énonce une échelle : 10^{-5} , négligeable à l'échelle humaine, 10^{-15} : négligeable à l'échelle terrestre, 10^{-50} : négligeable à l'échelle cosmique.

Paul Lévy, dans les années 1920 expose que le principe de Cournot est la seule façon d'établir un lien entre une théorie probabiliste et le monde extérieur aux mathématiques.

> *« La probabilité objective est tout à fait analogue à la longueur et à la masse, autres constantes physiques dont la signification empirique est également définie par des méthodes qui permettent de les mesurer avec un degré d'approximation raisonnable »* (Lévy, cité par Shaffer).

Le principe de Cournot (en fait appelé jusqu'en 1949 « loi fondamentale du hasard » ou « loi unique du hasard ») fut précisé en 1951 par Frechet qui distingua une forme faible et une forme forte du principe de Cournot :

• Pour un événement désigné avant une épreuve unique, si sa probabilité est faible ou nulle, l'événement ne se produira pas lors de cette épreuve unique ; voilà le principe fort.
• Un événement de probabilité très petite ne se produira que très rarement au cours d'épreuves répétées ; voilà le principe faible.
Shaffer (2007) écrit :

> *« Le principe fort se combine avec le théorème de Bernoulli pour produire la conclusion non équivoque selon laquelle la fréquence d'un événement sera proche de sa probabilité dans une série suffisamment longue d'épreuves indépendantes. Le principe faible se combine avec le théorème de Bernoulli pour donner la conclusion suivante : dans une série suffisamment longue d'épreuves indépendantes, la fréquence de l'événement sera* ***généralement*** *proche de sa probabilité, principe général dont le principe faible est un cas particulier. »*

Il demeure la question du sens que l'on donne dans la pratique à une probabilité considérée comme négligeable ; citons Martin [2003] :

« - Ou bien, on pose l'événement de très petite probabilité comme pratiquement impossible, parce qu'il est non pas seulement possible, mais surtout nécessaire ou préférable de n'en pas tenir compte dans la pratique. Ainsi, la prise de décision exige que l'on définisse les paramètres à mobiliser dans l'analyse des conséquences possibles du choix que l'on se propose d'effectuer ; et on s'autorise à négliger l'événement dont la probabilité est suffisamment faible, tout en sachant qu'il n'est pas, en toute rigueur, impossible. Tel est le cas lorsque l'action envisagée fait apparaître une disproportion marquée entre la probabilité très faible des risques encourus et les coûts élevés à mettre en œuvre pour les éviter. Ainsi compris, le principe des probabilités négligeables est un principe ***pragmatique*** *; il définit une probabilité* ***subjectivement négligeable****.*

- Ou bien, on pose l'événement comme physiquement impossible parce que, de fait, il ne peut se produire dans les conditions qui sont celles de l'expérience, ou n'a lieu que très rarement. L'événement est logiquement possible, mais il ne peut se produire parce que les conditions de son application au réel ajoutent des contraintes supplémentaires qui interdisent ou compromettent sa réalisation. Il s'agit cette fois d'un principe ***physique****, et il définit une probabilité* ***objectivement négligeable****. »*

La question posée ici n'est plus tout à fait celle de *probabilité objective* ou *probabilité subjective* telle que nous l'avons présentée au paragraphe 2.2. de ce chapitre, mais porte sur l'aspect *objectif* ou *subjectif* du caractère négligeable de la probabilité d'un événement dans un contexte donné, et d'une certaine manière in fine, de la nature du critère de la décision prise de tenir l'événement pour impossible : objective (n'importe qui prendrait la même décision) ou subjective (le décideur estime que compte tenu des autres paramètres liés au contexte, il peut tenir l'événement pour impossible).

Cette question de l'applicabilité des résultats du calcul des probabilités a été soulevée assez tôt par les mathématiciens, mais aussi par les philosophes, car elle interroge le lien entre probabilité et certitude.

Nous avons souhaité intégrer dans ce chapitre une référence au principe de Cournot et à la problématique d' « applicabilité » du calcul des probabilités dans la mesure où ces questions, peuvent peut-être trouver illustration dans l'enseignement secondaire grâce à la possibilité de répéter effectivement un grand nombre de fois une expérience simulée sur ordinateur, possibilité qui pourrait permettre d'illustrer par l'expérimentation l'observation (ou la non observation) des événements de faible ou de forte probabilité.

Mais surtout le principe de Cournot, qui n'est jamais évoqué dans les documents d'accompagnement à destination des enseignants, et probablement peu dans la formation des enseignants, nous paraît en réalité présent en filigrane dans les transpositions didactiques des programmes des deux dernières décennies : précisément dans la formulation qui est proposée de la loi des grands nombres.
Nous allons détailler ce point dans le chapitre suivant.

CHAPITRE 2

transpositions didactiques du calcul des probabilités

Jusqu'à présent, les transpositions didactiques du calcul des probabilités dans l'enseignement secondaire ne se sont jamais préoccupées de définir le hasard, ce qui n'empêche pas certains manuels d'en tenter des définitions, en préalable au cours de probabilités.
Les programmes se placent implicitement dans des situations de *hasard bénin* où la loi des grands nombres est valide.
L'objectif est d'introduire la notion de *probabilité d'un événement*, et les fondements du *calcul des probabilités* dans le cadre théorique de Kolmogorov, c'est à dire essentiellement la notion d'espace probabilisé modélisant l'expérience aléatoire à étudier.

Nous avons vu dans le chapitre précédent des enjeux épistémologiques autour de cette notion de *probabilité d'un événement*. Les programmes se placent implicitement dans le cadre des *probabilités objectives* : la probabilité d'un événement comme *mesure d'une propriété intrinsèque à la réalité*.
Nous avons décrit dans le chapitre précédent plusieurs démarches pour appréhender la notion de probabilité d'un événement donné : par une *approche géométrique* (ou *approche laplacienne*), par une *approche fréquentiste*.

Les programmes du lycée, pour introduire la notion de probabilité, se sont toujours appuyés sur des situations de *hasard du tirage au sort*, pas seulement parce qu'historiquement c'est dans ce cadre qu'a débuté le calcul des probabilités, mais surtout parce que la probabilité peut se calculer par dénombrement des cas favorables et des cas possibles, la loi de probabilité à associer à l'expérience étant uniforme : lancers de dés et lancers de pièces *équilibrés*, tirages *au hasard* d'une urne.
En France, jusqu'aux années 1990, l'enseignement des probabilités, quand il figurait dans les peogrammes était plutôt un prétexte à du calcul combinatoire : les notions de *permutations, p-listes, arrangements* et *combinaisons* étaient centrales.

1. Approche laplacienne et approche fréquentiste

1.1. L'approche laplacienne

L'approche laplacienne (ou géométrique) se place dans une situation où des considérations « évidentes » de symétrie(s) permettent de supposer l'équiprobabilité des issues. La probabilité de l'événement A considéré s'introduit « tout naturellement » [22] comme le rapport entre le nombre d'issues favorables à l'événement A et le nombre d'issues possibles à cette expérience, ce que les élèves retiennent sous la forme :

$$p(A) = \frac{\text{card}(A)}{\text{card}(E)} \quad \text{ou } p(A) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}.$$

Comment passer alors à une loi non équirépartie ?

Il faut bien à un certain moment définir la probabilité d'un événement sans l'hypothèse d'équiprobabilité des issues.

En fait, le cours classique, avec cette approche, consiste à introduire une application p de l'ensemble des événements vers [0 ; 1], ayant des propriétés « raisonnables » pour mesurer « correctement » le degré de possibilité que cet événement se produise. Ces propriétés « raisonnables » sont bien sûr :

(i) $p(E) = 1$ où E est l'événement certain

(ii) si A et B sont deux événements disjoints (« incompatibles »), $p(A \cup B) = p(A) + p(B)$

Ces propriétés paraissent d'autant plus « raisonnables » qu'on a en tête

$$p(\text{événement}) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}$$

puisque dans ce cas (i) est évidemment réalisé

et que (ii) renvoie à la propriété supposée « bien connue » des cardinaux :

si A et B sont deux ensembles disjoints, $\text{card}(A \cup B) = \text{card}(A) + \text{card}(B)$.

Donc le cours classique, dans cette approche, définit finalement l'espace probabilisé par des considérations purement ensemblistes, permettant les lois non équiréparties, mais avec en toile de fond l'hypothèse d'équirépartition des issues ; dans les périodes où les probabilités conditionnelles ne sont pas dans les programmes, les exercices comportent presque toujours

---

[22] Nous avons mis ici des guillemets pour souligner que rien de tout ceci n'est évident ni naturel, à cela près que les élèves qui reçoivent ce cours ont des connaissances spontanées qui en général s'accomodent assez bien de ce discours.

l'hypothèse d'équirépartition, exceptés pour quelques dés mystérieux dont la face 6 a 2 fois plus de chances de sortir que les autres.

Lorsque les probabilités conditionnelles sont dans les programmes, on peut avoir des situations où certaines probabilités sont données par l'« extérieur » : L'entreprise X fabrique des pièces A et B ; on sait que 10% des pièces A sont défectueuses etc. L'élève sait que ce 10% lui donne la probabilité d'un certain événement.
Il semble que cela fonctionne assez bien, surtout depuis que les arbres pondérés sont un outil de représentation, non seulement autorisé, mais préconisé. L'élève a-t-il une idée comment « on » sait que 10 % des pièces sont défectueuses ? Est-ce une question importante pour lui ?

1.2. L'approche fréquentiste

Comme nous l'avons dit, la conception implicite du hasard dans l'enseignement secondaire est le *hasard bénin*, pour lequel la loi des grands nombres est valide.
On s'occupe donc d'expériences aléatoires reproductibles, les épreuves étant indépendantes : lancer de pièces, de dés, tirages d'une urne avec remise, etc.
Pour une telle expérience E que l'on répète *n* fois, on s'intéresse pour chaque épreuve à la réalisation ou non d'un événement A lié à l'expérience.
Les élèves doivent constater, ou admettre, que si *n* prend des valeurs de plus en plus grandes, les fréquences f_n de l'événement A tendent à se stabiliser vers un nombre « théorique » compris entre 0 et 1, qu'on appelle la probabilité p(A) de l'événement A.
On définit ainsi (au moins en théorie) une application p de l'ensemble des événements vers l'intervalle [0 ; 1].

La notion d'espace probabilisable [23] lié à l'expérience aléatoire, et surtout la terminologie : *issue (ou événement élémentaire), événement, événement contraire*, etc. peut se mettre en place à partir de l'étude des fréquences de réalisation des événements liés à une expérience aléatoire.
On constate les propriétés des fréquences des événements lorsqu'on répète *n* fois une expérience : si f_n associe à un événement sa fréquence d'apparition lors de ces *n* répétitions, la

[23] C'est à dire : un univers U et un ensemble d'événements qui est l'ensemble des parties de U dans toutes les prescriptions exceptée les programmes « des maths modernes ».

fonction f_n vérifie entre autres propriétés

(i) $f_n(U) = 1$ où U est l'événement certain

(ii) si A et B sont disjoints, $f_n(A \cup B) = f_n(A) + f_n(B)$

Il est alors « naturel » de penser que « à la limite » [24] l'application probabilité p vérifie :

(1) $p(U) = 1$

(2) si A et B sont disjoints, $p(A \cup B) = p(A) + p(B)$

On peut alors remarquer que l'application p est parfaitement déterminée dès que l'on connaît les images par p de toutes les issues.

1.3. Faut-il privilégier l'approche fréquentiste ou l'approche laplacienne ?

Ces deux approches supposent une même structure de départ : un espace probabilisable (Ω, B) fini, où B est une tribu sur l'ensemble des parties de Ω, l'univers Ω étant supposé fini, (sauf pour une partie des programmes de TS et TES de 2003) ; passée l'époque des « mathématiques modernes », B est tout simplement P(Ω), l'ensemble des parties de Ω .

Les deux approches débouchent sur une même structure : un espace probabilisé fini (Ω, B, p) où p est une application probabilité définie sur B.

L'approche laplacienne risque, peut-être, de renforcer le biais d'équiprobabilité des issues, (nous exposons dans ce chapitre au paragraphe 4 ce biais), biais qui fait facilement partie des connaissances spontanées des élèves, alors qu'il s'agit justement, au delà des connaissances spontanées, d'introduire une théorie qui prend en compte des situations « non géométriques », c'est à dire ne relevant pas d'une loi uniforme.

L'approche fréquentiste pose la question du statut de la loi des grands nombres, plus exactement, pour ce qui concerne la transposition de la théorie des probabilités dans l'enseignement secondaire : la *stabilisation des fréquences* des issues, que l'on peut observer en répétant un grand nombre de fois l'expérience étudiée.

En résumé, la différence des approches réside donc dans la manière d'introduire l'application probabilité.

[24] Nous avons mis des guillemets pour souligner à nouveau que ceci n'est peut-être pas « naturel » : est ici en jeu non seulement la définition de la probabilité comme limite d'une suite qui change à chaque nouvelle série de répétitions *n* fois de l'expérience, mais également une conservation d'égalité « par passage à la limite ».

Citons Parzysz (2003) , p. 33 :

> *« [Les deux approches laplaciennes et fréquentistes] apparaissent donc finalement '' contradictoires et complémentaires'', et, en conséquence, se limiter à l'une ou l'autre des deux approches est, comme on l'a vu, nécessairement réducteur. C'est pourquoi une démarche didactique cohérente se devrait d'intégrer les deux aspects, en prenant en compte les conceptions ''initiales'' des élèves et les obstacles épistémologiques repérés. »*

Les approches sont contradictoires en ce sens que l'approche laplacienne est purement théorique, liée à « la géométrie » du hasard, et limitée à des situations très particulières où la loi uniforme s'impose, alors que l'approche fréquentiste paraît d'essence expérimentale, mais adaptée à toute situation (relevant du *hasard bénin*).

Elles sont complémentaires, d'une part parce la démarche fréquentiste ne prend sens qu'en se référant d'abord aux situations simples relevant d'une loi uniforme, en particulier les situations où les élèves ont des conceptions spontanées , et d'autre part parce que l'approche fréquentiste peut permettre de valider (ou non) le modèle qu'une approche laplacienne aura construit, mais aussi permet d'approcher une loi de probabilité qui ne relèverait pas en dernier ressort d'une loi uniforme (c'est à dire ne serait pas image d'une loi uniforme par une variable aléatoire).

2. Le statut de la loi des grands nombres dans les programmes

Quel est le statut de la loi des grands nombres dans les programmes ? S'agit-il uniquement d'un théorème interne à la théorie des probabilités, qui dit quelque chose sur la loi de probabilité de l'expérience considérée, <u>à l'intérieur du modèle choisi</u> ? S'agit-il d'un *principe (fort* ou *faible*) basé sur le principe de Cournot ? Ou encore s'agit-il d'une propriété totalement objective de la réalité de cette expérience ?

Cette question n'a pas à être présentée en tant que telle dans la transposition didactique (en lycée) de la théorie des probabilités, mais elle est en filigrane dans certains débats et essais qui ont accompagné, par exemple dans les IREM, les programmes des années 2000.

Par ailleurs le choix de l'énoncé de la loi des grands nombres que la prescription propose à un moment donné, est peut-être lié au statut assigné, par les prescripteurs, à cette loi des grands nombres.

Nous allons illustrer cela en reprenant les évocations de la loi des grands nombres qui apparaissent dans les programmes (source : une compilation de Bernard Courtebras pour un séminaire de l'EHSS : « séminaire d'histoire du calcul des probabilités et de la statistique »). Tous ces programmes sont publiés aux B.O.E.N. [25] et consultables en ligne.

Signalons que de 1986 à 1992, l'enseignement des probabilités débute en classe terminale.

période	niveau	évocation de la loi des grands nombres [26]
de 1986 à 1992	Terminales A1, A2, A3, B, C et E	*(...) on mettra en évidence la <u>stabilité</u> de la fréquence d'un événement donné lorsque l'expérience est répétée un grand nombre de fois.*
de 1991 à 2001 de 1991 à 1992	Premières A1, A2, A3, S et E Premières B	*(...) la <u>relative stabilité</u> de la fréquence d'un événement donné lorsque l'expérience est répétée un grand nombre de fois.*
de 1993 à 2001	Premières ES (anciennes B)	*(...)On pourra partir de la répétition d'une expérience aléatoire et introduire la notion de probabilité en soulignant les propriétés des fréquences et la <u>relative stabilité</u> de la fréquence d'un événement donné lorsqu'on répète l'expérience un grand nombre de fois. Dans ce cas on s'appuiera sur une notion intuitive et non formalisée de l'idée de répétition <u>sans poser le problème de l'indépendance des expériences</u> et on procédera surtout par travaux pratiques (utilisation de calculatrices, lancers de pièces...). »*
de 1994 à 1998	Terminales ES	*(...) c) Loi des grands nombres : Dans le cas du schéma de Bernouilli, mise en évidence expérimentale de la <u>convergence</u> de la fréquence empirique du nombre de succès vers la probabilité théorique. [La formulation mathématique de la loi des grands nombres est hors programme. On s'appuiera sur des expériences aléatoires (tirage dans une urne, <u>simulation sur ordinateur</u> [27] ...)]*

[25] Bulletin Officiel de l'Education Nationale.

[26] Tout ce qui est souligné dans cette colonne l'a été par nous.

[27] souligné par nous. Il semble que c'est la première fois que l'idée de simulation sur ordinateur apparaît dans la prescription.

On remarquera une forme de progression dans ces énoncés : *stabilité*, puis *relative stabilité*, puis *relative stabilité* toujours, mais en signalant qu'il y a des hypothèses (l'indépendance des essais), même si ces hypothèses ne doivent pas être formalisées avec les élèves.
L'expression *convergence* apparaît en 1994 dans les programmes de terminale ES.
Des expressions telles que *stabilité* ou *convergence* occultent le fait qu'il s'agit seulement d'une convergence en probabilité, ce qui pourrait illustrer le *principe de Cournot*.
L'ajout de l'épithète *relative* fait en revanche réapparaître la convergence en probabilité.

A partie de 1991, les probabilités sont introduites dès la classe de première, alors que l'analyse combinatoire relève de la classe terminale. Cela a pour effet de faciliter (au moins théoriquement) l'approche fréquentiste en déconnectant le calcul des probabilités de l'analyse combinatoire : le dénombrement des cas favorables et des cas possibles, dans le cas d'une loi équirépartie, ne peut plus se faire que sur des exemples simples.

Pour conclure, pour cette période de 1986 à 2001 :
1. Les programmes préconisent une approche fréquentiste basée sur une convergence des fréquences d'un événement vers sa probabilité, convergence énoncée d'abord en terme de *stabilité* des fréquences, puis de *relative stabilité*, puis de *relative stabilité sous certaines hypothèse pour l'expérience* lorsqu'on répète un grand nombre de fois l'expérience.
2. Les probabilités ne sont plus présentées comme un appendice de la combinatoire, un prétexte à dénombrement, mais comme une partie spécifique des mathématiques.
3. La part du calcul combinatoire diminue considérablement, se limitant pour les « programmes 2000 » aux seules *combinaisons*.
4 . En corollaire, il est fermement affirmé que le calcul des probabilités ne doit pas amener de difficultés techniques de dénombrement.
On a en quelque sorte un renversement des rôles : la combinatoire n'est plus qu'un outil pour le calcul des probabilités alors que les probabilités n'étaient auparavant qu'un prétexte à dénombrement.
5. Il est mentionné, pour la première fois, en terme encore fort vague, la possibilité de faire des *simulations sur ordinateur*, et la possibilité d'*expérimentations* pour illustrer la loi des grands nombres (programme de terminale ES 1994-1998).

Les programmes des années suivantes (2000) introduisent explicitement la loi des grands nombres par un énoncé « vulgarisé ». Les programmes de première S par exemple stipulent :

« *Le lien entre loi de probabilité et distributions de fréquences sera éclairé par un énoncé vulgarisé de la loi des grands nombres.* »

Ces programmes proposent l'énoncé :

« *Pour une expérience donnée, dans le modèle défini par une loi de probabilité P, les distributions des fréquences calculées sur des séries de taille n se rapprochent de P quand n devient grand* »

Programmes de première S, in B.O. n°7 Hors-série, 31 août 2000.

Et on trouve dans les documents d'accompagnement :

« *Dans le monde théorique défini par une loi de probabilité P sur un ensemble fini, les fréquences des éléments de cet ensemble dans une suite de n expériences identiques et indépendantes «tendent» vers leur probabilité quand n augmente indéfiniment.* »

Ces énoncés semblent proposer une convergence certaine, avec le choix du mode indicatif : « se rapprochent de », ou « tendent », ce qui nous renvoie plutôt à la forme forte du principe de Cournot ; on aurait pu trouver un mode conditionnel : « devrait tendre vers ou se rapprocher de » , ou une expression de réserves : « est généralement proche de », ou « est relativement proche », , ce qui renverrait plutôt à la forme faible de ce principe.

Mais finalement le deuxième énoncé, du fait des guillemets qui encadrent ce *tendent* , laisse la libre interprétation : forme faible ou forme forte du principe de Cournot, mais ces guillemets peuvent tout aussi bien marquer seulement le fait qu'il ne s'agit pas de la convergence certaine des suites numériques du cours d'analyse. Il n'y aurait alors aucune référence à l'interprétation pratique (l'*applicabilité* des probabilités) que l'on peut vouloir faire de la loi des grands nombres.

Par ailleurs, nous comprenons de la façon suivante la première partie de cet énoncé (Programmes de première S, in B.O. n°7 Hors-série, 31 août 2000) : la loi de probabilité P associée à une expérience est d'ordre théorique, mais pas simplement en terme de « limite idéale » jamais atteinte comme la limite 0 pour la suite numérique de terme général (1/n), mais plus fondamentalement parce qu'elle correspond au choix d'un modèle (éventuellement parmi d'autres possibles) ; et c'est dans le cadre de ce modèle qu'il y a convergence des fréquences des issues vers leur probabilité. C'est peut-être une manière de rappeler au lecteur averti qu'il s'agit en fait d'une convergence en probabilité.

Nos observations spontanées dans nos classe nous forcent à constater que ces deux énoncés n'évoquent pas grand chose auprès des élèves. Ces énoncés semblent plutôt destinés à la communauté scientifique, comme présentation d'une loi des grands nombres vulgarisée certes, mais le lecteur averti peut l'interpréter comme une « traduction » de l'énoncé : quel que soit l'écart $e > 0$, l'événement « l'écart absolu entre f_n et p dépasse e » a une probabilité qui tend vers zéro lorsque n tend vers $+ \infty$.

Nous verrons plus loin, dans le chapitre 3, que la possibilité de simulations en grand nombre d'expériences aléatoires rendait probablement nécessaire la présentation d'un énoncé de la loi des grands nombres moins vague que simplement l'énoncé d'une *relative stabilité* des fréquences, en particulier l'expression d'un lien avec le modèle choisi, surtout dans le cadre d'une prescription qui aborde la validation du modèle choisi, comme c'est le cas avec les « programmes 2000 ».

3. Les finalités d'un enseignement des probabilités

L'analyse proposée en 1999 par Lahanier-Reuter, en quelque sorte l'état des lieux de l'enseignement des probabilités dans l'enseignement secondaire, semble avoir conservé en grande partie son actualité. Nous allons donc présenter cette analyse, puis décrirons quelques « avancées » dans la prescription.

3.1. L'analyse de Lahanier-Reuter

A propos d'une étude des exercices donnés au bac (session 1994), Lahanier-Reuter écrit (p. 59) :

> *« L'intérêt de la modélisation de la situation est à reconstruire par l'élève, puisque la synthèse des informations nouvelles que permet cette modélisation n'est jamais ni proposée, ni demandée. »*

Et plus loin, p. 62 et 63 :

> *« En conclusion, le hasard est mobilisé dans les exercices du baccalauréat dans des situations qui sont sans exceptions des situations de modélisation. Les conditions de modélisation diffèrent : elles peuvent être de consensus, dans le cas des jeux de hasard, testées, éprouvées, dans le cas des situations de stratégie, hypothèses d'expérimentation dans le cas d'expérience scientifique.*
>
> *Mais si les questions posées semblent cohérentes par rapport à cette*

classification, il faut souligner que les fonctions modélisatrices du hasard ne sont jamais explicitement décrites, et que par conséquent elles sont à la charge de l'élève. De même les questions s'enchaînent, des calculs sont demandés, des informations nouvelles sont produites, sans que les auteurs de ces exercices ne demandent, ne présentent ou n'exigent un retour de ces productions sur la situation initiale : la fonction du recours au hasard, à l'aléatoire ne reste-t-elle pas alors entièrement à construire par les élèves, et peut-on prendre le risque d'affirmer qu'elle l'est ? ».

Lahanier-Reuter conclut son chapitre "Le hasard, objet de l'enseignement mathématique" par une hypothèse concernant ce qu'elle nomme "*les instants choisis de la découverte*" : situations à caractère plutôt expérimentales pour des enfants d'âge entre 9 et 12 ans, situations qui ont fait l'objet d'expérimentations, fructueuses d'après leurs auteurs, ou apprentissage des situations aléatoires en classes de première et terminale de lycée ?
Dans le premier cas, l'objectif serait de « *lutter contre l'intuition commune, la superstition* », et dans le deuxième cas l'objectif semblerait plutôt :

> *« un enseignement préalable à un enseignement théorique des probabilités. Il est constitué de définitions "élémentaires" des "concepts de base" : les ensembles sont finis, les variables aléatoires sont discrètes, les lois de probabilité réduites à la loi binomiale, etc. <u>Ajoutons d'après l'étude des exercices menée plus haut, que cet enseignement a pour but essentiel de faire fonctionner les outils probabilistes définis dans des calculs bien précis, bien plus que de donner un sens à ce qui constitue une problématique probabiliste</u>.[28]*
> *Cet enseignement des probabilités, de par les caractéristiques décrites, paraît bien davantage être conçu en tant qu'enseignement "pré-universitaire" d'un savoir théorique, scientifique encore inaccessible. ».*

Une initiation précoce serait l'expression « *d'une conception de la fonction sociale, en particulier civique, que revêt cet enseignement* ».
Une initiation en fin de lycée correspondrait à une fonction sociale de cet enseignement de l'aléatoire « *qui inscrit ces connaissances dans un cursus* » :

> *« En conclusion, deux conceptions divergentes sur l'enseignement de l'aléatoire,*

[28] Passage souligné par nous.

conceptions qui sont traces de ce que peuvent être les fonctions sociales d'un tel enseignement, peuvent être isolées : soit un enseignement utile en ce que les apprentissages visés seront des apprentissages utiles à détourner, à affaiblir la force des connaissances "vulgaires", considérées comme dangereuses socialement (pratiques des jeux de hasard, croyances et superstitions), soit un enseignement qui envisage ces apprentissages comme des "gammes" nécessaires, comme d'indispensables préliminaires à tout enseignement supérieur (universitaire) théorique ».

3.2. Des avancées depuis 1999 ?

Les programmes de mathématiques du lycée et du collège ont changé dans les années 2000, puis en 2010. Nous présenterons dans le chapitre 3 ce qui concerne l'enseignement des probabilités, et tout particulièrement les simulations d'expériences aléatoires.
Les nouveaux programmes (2000 et 2010) semblent aller dans le sens souhaité par Lahanier-Reuter : des lois continues sont introduites en Terminale scientifique dans les programmes de 2001 ; la notion de probabilité d'un événement est introduite dès la classe de troisième ; des finalités sociales pour un enseignement de la statistique et des probabilités sont explicitement formulées dans ces programmes et les documents qui les accompagnent.

Les arguments pour justifier une part plus importante de l'enseignement de la statistique et des probabilités dans les programmes des années 2000 (voir le chapitre 3) font explicitement référence aux besoins sociaux .

Ces programmes introduisent en terminale des lois de probabilité continues (lois uniformes sur un segment de IR, loi exponentielle sur IR^+, modélisant une loi à durée de vie sans vieillissement (la démonstration de l'équivalence entre les deux lois est du programme de terminale S). La loi exponentielle doit être étudiée simultanément en classe de mathématiques et en classe de sciences physiques dans le contexte de la désintégration des noyaux.

Moins d'une décennie plus tard (rentrée 2008), un enseignement des probabilités est introduit dans les programmes de troisième (âge autour de 14-15 ans), enseignement poursuivi en seconde (rentrée 2009).

Il demeure cependant bien des réticences dans le corps enseignant, tant sur l'importance de l'enseignement des probabilités dans les programmes, que sur l'avancée dans le cursus scolaire (classe de troisième) de cet enseignement.

Ces réticences s'expriment sous diverses formes :

• par des prises de position, parfois véhémentes, sur les forums de discussion, par exemple de l'A.P.M.E.P. [29] ,

• par une certaine inertie dans la mise en œuvre de la prescription.

Dans une contribution à un bulletin de l'A.P.M.E.P. , Didier Dacunha-Castelle (2009) écrit :

> *« Les traditions nationales dictent en partie les contenus des programmes. Cela vaut pour les mathématiques et les probabilités en sont l'exemple. Il est temps de se demander pourquoi l'on en est encore aujourd'hui à discuter de l'introduction de l'aléatoire dans les programmes du collège et du lycée, avec un retard singulier par rapport à bien des pays. Un blocage culturel a fait place à un blocage institutionnel.*
>
> *(...)*
>
> *[en] 2009 : Les classes préparatoires M et M' restent dans le monde le seul premier cycle mathématique sans enseignement de l'aléatoire. Les programmes de physique ressemblent étrangement à ceux des années 60. Sans introduction des probabilités ils ne peuvent être modernisés. Le problème de l'enseignement universitaire est aujourd'hui réglé y compris en mathématiques où cela fut le plus difficile !*
>
> *Pour le secondaire, les programmes du lycée sont à transformer doucement surtout en S. L'introduction de l'aléatoire dès le collège est le sujet des articles détaillés de ce bulletin.*
>
> *Je résumerai ainsi les arguments qui militent en faveur de cette introduction :*
>
> *- Les probabilités sont comme la géométrie à connotation expérimentale.*
>
> *- Fausses et vraies intuitions s'entremêlent dès le plus jeune âge, ce qui donne beaucoup de sens à la recherche de la rigueur dans le raisonnement.*
>
> *- Les probabilités permettent de travailler la logique élémentaire du ou, et, si (surtout le si !) ... abandonnée en rase campagne avec les ensembles lors du rejet de la réforme des « maths modernes », abandon détestable dont les effets*

---

[29] Association des Professeurs de Mathématiques de l'Enseignement Public.

perdurent. Le raisonnement probabiliste à un niveau élémentaire s'appuie sur des arbres, des graphes, exercices de logique élémentaire.

- Les probas-stats sont par excellence le domaine où le dictionnaire langue quotidienne / langage des mathématiques fonctionne à plein régime, avec un vrai travail sur la langue pour l'élève dont l'importance est considérable.

- Les probas-stats sont enseignées aujourd'hui dans tous les premiers cycles non littéraires et donnent hors enseignement les principaux débouchés de nos étudiants. Cela touche toutes les sciences et d'abord le cœur de la biologie, la technologie (signal, image, télécom), l'industrie (contrôle de qualité, ...), le tertiaire et la gestion (files d'attente, stocks, assurance) et bien sûr l'économie.

- L'aléatoire est le concept mathématique clé pour comprendre les grandes théories qui irriguent la pensée contemporaine et d'abord la théorie de l'évolution, où le hasard est au cœur des phénomènes de sélection, la génomique, et bien sûr les mécaniques quantique et statistique.

• En avançant du primaire à la sortie de l'Université on voit que les probas-stats sont partie prenante d'un grand nombre d'activités humaines. L'absence de compréhension de l'incertain est une des causes, pas la seule, de l'irrationalité bien au-delà de la pratique des jeux de hasard.

La compréhension des concepts liés aux risques : écologique, climatique, médical et épidémiologique doit être partagée largement car c'est une clé pour la décision politique dans bien des domaines. Le triste et grave épisode de la crise financière où une partie de notre communauté a, hélas, une responsabilité certaine, ne fait que renforcer la nécessité de donner à tous des outils pour analyser le risque. »

4. Des difficultés didactiques répertoriées par les chercheurs

4.1. Les biais

Nous renvoyons à la thèse de Nabbout (2006), p.18 et 19, pour un exposé de sept *biais du raisonnement* décrits par divers chercheurs.

Nous présentons ici les biais que nous avons rencontrés dans la partie expérimentale de notre thèse.

4.1.1. Le biais d'équiprobabilité

Cette difficulté, ce peut être un obstacle, consiste à associer, au moins dans une étape spontanée, la loi uniforme aux issues d'une expérience aléatoire : par exemple une probabilité de 1/11 à chacune des issues 2, 3, …, 12 obtenues en lançant deux dés cubiques et en s'intéressant à la somme des points amenés.

4.1.2. Le biais de représentativité et la loi des petits nombres

> *« La loi des grands nombres garantit que les échantillons très grands seront en effet fortement représentatifs de la population dont ils sont tirés. En outre, si une tendance auto-correctrice s'opère, alors les petits échantillons devraient aussi être fortement représentatifs et semblables entre eux. Les intuitions des gens en ce qui concerne les échantillons aléatoires semblent satisfaire à la loi des petits nombres, prétendant que la loi des grands nombres s'applique aussi aux petits nombres. »* Kahneman et Tversky (1972)

La *loi des petits nombres* peut être considérée comme un cas particulier du *biais de représentativité* : dans certaines situations la décision d'un individu peut être fondée sur *l'heuristique de représentativité* ; c'est le cas d'une majorité des personnes interrogées par Kanheman et Tarsky lors de l'expérience "le portrait de Linda" : ces personnes ont confondu l'événement qui leur paraissait le plus représentatif de la description qui leur avait été faite de Linda, avec l'événement le plus probable. Cette heuristique a alors amené une *intuition erronée* que l'on peut formaliser par *un effet de conjonction*, selon lequel si E et F sont deux événements, la probabilité $p(E \cap F)$ serait supérieure à $p(E)$.

De cette *heuristique de représentativité* découle un autre type d'intuition erronée : *la loi des petits nombres*, qui amène à dire que si, ayant lancé 3 fois une pièce de monnaie, on a obtenu Face, Pile et Pile, alors au quatrième lancer, il est plus probable que ce soit Face qui sorte, plutôt que Pile ; pour rendre compte de l'équiprobabilité des faces, la séquence FPPF serait plus représentative que la séquence FPPP : la pièce étant supposée équilibrée, Pile a autant de chance de sortir que Face, donc sur quatre lancers, observer autant de Piles que de Faces représenterait mieux cette égalité des chances qu'un excès de Piles.

On parle de loi des petits nombres ; l'expression renvoie bien sûr à la loi des grands nombres, en jouant sur deux facettes :

- de fait, lorsqu'on répète un grand nombre de fois une expérience aléatoire, la distribution des fréquences observées converge (en probabilité) vers la loi de probabilité, donc les effectifs

observés des différentes issues sont *représentatifs* de cette loi de probabilité, c'est à dire à peu près dans les mêmes rapports que les probabilités de ces issues ;
- Cette *représentativité* devrait s'observer lorsque l'on répète quelques fois seulement l'expérience (un petit nombre de fois).

Il n'est pas sûr que la genèse de ce biais de *la loi des petits nombres* passe par la loi des grands nombres : si l'issue e_1 a 2 chances sur 7 d'être observée lorsqu'on réalise une certaine expérience aléatoire, l'issue e_2 4 chances sur 7 et l'issue e_3 1 chance sur 7, l'heuristique de représentativité peut amener à croire que l'apparition de chacune des issues e_1, e_2 et e_3 devrait se faire proportionnellement avec les nombres 2, 4 et 1, quel que soit le nombre de répétitions de l'expérience.
Comme il s'agit d'expériences aléatoires, la proportionnalité n'est pas respectée à l'issue de chaque essai : il y a du hasard, mais un nouvel essai devrait tendre à corriger les « déséquilibres », à défaut de ramener l'équilibre comme dans l'exemple du lancer d'une pièce décrit plus haut.
Ce biais de représentativité explique certaines réponses à des questions telles que : « on a lancé déjà 5 fois le dé, et on a sorti 6, 4, 6, 6, 2 ; on lance une sixième fois le dé, qu'est-ce qui est le plus probable : sortir un 6 ou un 5 ? ». Les réponses en faveur du 5 permettent de corriger le déséquilibre en faveur du 6.
Nous analyserons dans la troisième partie certaines réponses d'élèves comme relevant de ce biais de *la loi des petits nombres*.

4.2. Des problématiques nouvelles

D'autres difficultés paraissent plus intrinsèques à l'enseignement de la théorie du calcul des probabilités chez l'élève, ce qui a pu amener des expressions telles que « former à l'esprit statistique ».
Ces difficultés sont pointées par Lahanier-Reuter (1999), p.48 :

> *« Surgissent des outils nouveaux de production de connaissances ainsi que des connaissances inaccessibles par les outils traditionnels. Par exemple, le hasard lié à l'épreuve aléatoire permet de construire une théorie de l'échantillonnage. Regard neuf sur d'anciens phénomènes ou sur des phénomènes inexplorés, accessibles à des outils théoriques nouveaux, les problématiques dans lesquelles le hasard s'inscrit dans l'enseignement statistique et probabiliste sont essentiellement nouvelles et en rupture par rapport aux problématiques déjà*

abordées. »

Ainsi les rapports au *certain* et au *déterminisme* ont beau être contournés par des exercices tels que ceux décrits par Lahanier-Reuter, et peut-être encore plus, récemment, avec une utilisation quasi mécanique des arbres de probabilité, des questionnements demeurent auprès des élèves, parfois cachés derrière des questions telles que « combien de fois faut-il répéter la simulation ? », sans que la finalité soit précisée : « combien de fois faut-il répéter la simulation pour que … ? ».
Nous verrons ainsi dans l'analyse de notre expérimentation (troisième partie) que pour certains élèves, répéter suffisamment la simulation donne une certitude.

4.3. Les opérations

Lahanier-Reuter cite, parmi les difficultés, « *l'opération de substitution sur le statut des nombres rationnels* » rencontrés lors du calcul des probabilités, qu'ils soient des fréquences (nombre d'issues observées sur nombre d'épreuves), ou résultats d'un « calcul laplacien » : nombre de cas favorable sur nombre de cas possible ; ces nombres passent de la qualité de nombres rationnels, quotients de deux entiers issus d'opérations concrètes de dénombrement, à la qualité de nombres réels de l'intervalle [0 ;1], mesures de la probabilité d'un événement.
Nous avons trouvé dans notre expérimentation quelques réponses où en lieu et place des probabilités attendues dans le tableau de la loi de probabilité, figuraient le nombre de cas favorables à chaque issue, le nombre de cas possibles apparaissant par ailleurs dans le travail rendu.

Lahanier-Reuter remarque que cette opération de substitution sur le statut de ces nombres se fait dans les deux sens : *dans l'étape de modélisation* la probabilité p d'un événement « passe » de IQ à IR, et « dans un mouvement inverse » dans *l'étape de simulation*, p « passe » de IR à IQ.
Notons que ce mouvement inverse lors de la simulation d'une expérience aléatoire a lieu lorsque la simulation consiste, par quelque procédé que ce soit, à simuler la loi de probabilité associée à l'expérience par un tirage d'urne :
si l'expérience admet k issues $x_1, x_2, \ldots, x_k$ avec les probabilités $p_1, p_2, \ldots, p_k$, les p_i étant des réels rationnels (par exemple donnés sous forme décimale), on met les p_i au dénominateur commun b : $p_i = \frac{a_i}{b}$ où les a_i et b sont des entiers, et on simule l'expérience par un tirage d'un

jetons d'une urne contenant b boules de k couleurs différentes, l'effectif de la $i^{ème}$ couleur étant a_i. Bien entendu, $\sum a_i = b$.

4.4. Une nouvelle forme de raisonnement en mathématiques

Lahanier-Reuter cite également, parmi les difficultés, « *la déduction par induction dans la statistique inférentielle* » : cette démarche n'apparaît pas explicitement dans la transposition actuelle pour les lycées telle qu'elle se présente dans les programmes des années 2000, mais elle se profile dans les programmes des années 2010 (cf. chapitre 3). Notons qu'elle est en fait implicitement omniprésente : par exemple, lorsque dans un exercice il est énoncé « on a constaté que la probabilité qu'une pièce fabriquée par l'usine A soit défectueuse est égale à 0, 07 ». De tels énoncés sont courants, particulièrement lorsque leurs auteurs veulent sortir des sentiers battus du hasard des jeux.

4.5. La validité d'un raisonnement

Lahanier-Reuter souligne le déplacement des propositions sur lesquelles s'exercent « *dans ce système de pensée* » la validation Vrai/Faux : ce n'est pas sur le fait que le joueur a fait un double six en lançant les deux dés que s'exerce cette validation Vrai/Faux, mais sur le fait que la probabilité de sortir un double six est de 1/36.

Nous allons préciser ce point : la validation ne porte pas particulièrement sur la probabilité égale à 1/36 de sortir le double six, résultat d'un calcul mené sous l'hypothèse que les dés sont parfaitement équilibrés. Sous une autre hypothèse, ce 1/36 pourra être faux.

Lahanier-Reuter a écrit :

> *« Dans ce système de pensée, ce sont désormais les mesures en probabilité des événements qui sont décrétées vraies ou fausses, qui sont soumises à la vérification*[30]*. »*

Il ne s'agit donc pas du résultat 1/36 en lui même, qui est le résultat d'un calcul exact ou inexact dans un cadre de raisonnement et de calcul donné classique, lorsqu'a été choisi l'espace probabilisé avec lequel on a modélisé l'expérience aléatoire.

La validation porte précisément sur ce choix de cet espace probabilisé.

Cette validation peut se faire, dans le cadre du hasard bénin, donc dans le cadre du hasard

[30] Souligné par nous.

envisagé dans la transposition didactique au lycée, par la confrontation des résultats du calcul à des résultats empiriques obtenus en répétant un très grand nombre de fois l'expérience, en déterminant les fréquences observées pour chaque issue.

Lorsqu'il est possible d'obtenir de tels résultats empiriques, se pose la question : combien d'épreuves sont nécessaires pour valider ou invalider le choix ? [31] « Un très grand nombre de fois », est une notion bien relative.
Les propositions des élèves (de 10 à plusieurs milliers) lors de notre expérimentation, révèlent une certaine perplexité, que ne résout pas l'argument : « plus on répète, plus c'est précis ».
La théorie, certes, dit des choses là dessus, mais les élèves concernés, justement, ne connaissent pas encore cette théorie.

Est-ce que des « règles », même énoncées avec confrontation avec des données obtenues dans une situation ou deux comme cela est proposé dans les programmes 2010 (voir chapitre 3), résoudront cette difficulté ?

Cette rupture dans le fonctionnement des mathématiques n'est pas spécifique au calcul des probabilités : on la retrouve en logique (logique classique ou logique sans le tiers exclus par exemple), en géométrie (géométrie euclidienne ou géométrie non euclidienne par exemple).
Ce n'est pas non plus spécifique aux mathématiques : mécanique newtonienne ou mécanique relativiste en sciences physiques par exemple.

Mais dans l'enseignement des mathématiques, dans la scolarité secondaire, l'enseignement des probabilités est le seul lieu où apparaît (ou *devrait* apparaître, ou *pourrait* apparaître) le fait que la validité du modèle dépend du choix des « axiomes » (« axiomes » locaux, liés à l'expérience étudiée), et que se pose la question de leur validation.
Et cette situation nouvelle heurte l'idée que la validation des résultats en mathématiques ne se ferait que par la seule « *démonstration mathématique* » qui s'appuie sur la règle du *modus ponens*, démonstration qui est la norme visée par les programmes de mathématiques à partir

---

[31] On peut envisager d'introduire la notion de tests d'hypothèses dans l'enseignement secondaire (c'est le cas en terminale dans les programmes 2000), mais probablement pas au niveau du collège, et, par rapport aux objectifs visés dans un enseignement « rationnel » des statistiques-probabilités, probablement pas sous la forme d'un test « boîte noire » tel que peut être enseigné le test du chi-2 dans certaines études post-bac.

de la classe quatrième.

Là réside peut-être l'une des causes des réticences exprimées de façons variées par le corps enseignant vis à vis de l'enseignement des probabilités, parfois même en contestant l'appartenance du calcul des probabilités aux mathématiques [32] : réticences à relativiser les résultats démontrés en les re-situant dans un choix de système d' « axiomes ».
Lahanier-Reuter (1999) conclut (p. 50) :

> « *C'est donc un nouveau cadre épistémologique qui est proposé aux élèves et aux étudiants.*»

5. Une difficulté peu évoquée : les différents types de convergence

Comme nous l'avons déjà souligné, l'énoncé de la loi des grands nombres, quels que soient les termes choisis, évoque la convergence de la fréquence f_n d'un événement (lorsqu'on a répété *n* fois l'expérience) vers sa probabilité.
Il s'agit d'une *convergence en probabilité* :
quel que soit $\varepsilon > 0$, la limite, quand *n* tend vers $+\infty$, de la probabilité que l'écart absolu entre la fréquence f_n d'un événement et sa probabilité p dépasse ε est nulle,
pour l'énoncé de la loi « faible ».
La probabilité que (f_n) converge vers p est égale à 1, pour l'énoncé de la loi « forte ».

Cela ne signifie pas la même chose qu'une convergence « certaine » de la suite (f_n), telle qu'elle est définie en analyse.
Or cette convergence « certaine » des suites numériques figure au programme de mathématiques de première S.
La même année de première S donc, les élèves sont confrontés à deux types de convergence pour des suites numériques.
Les programmes et les documents qui les accompagnent n'ont pas jusqu'à présent abordé frontalement cette difficulté potentielle.

Les programmes de première S proposent l'énoncé :

> « *Pour une expérience donnée, dans le modèle défini par une loi de probabilité P, les distributions des fréquences calculées sur des séries de taille n se rapprochent*

[32] dans les forums de discussion sur le site de l'APMEP par exemple.

de P quand n devient grand »

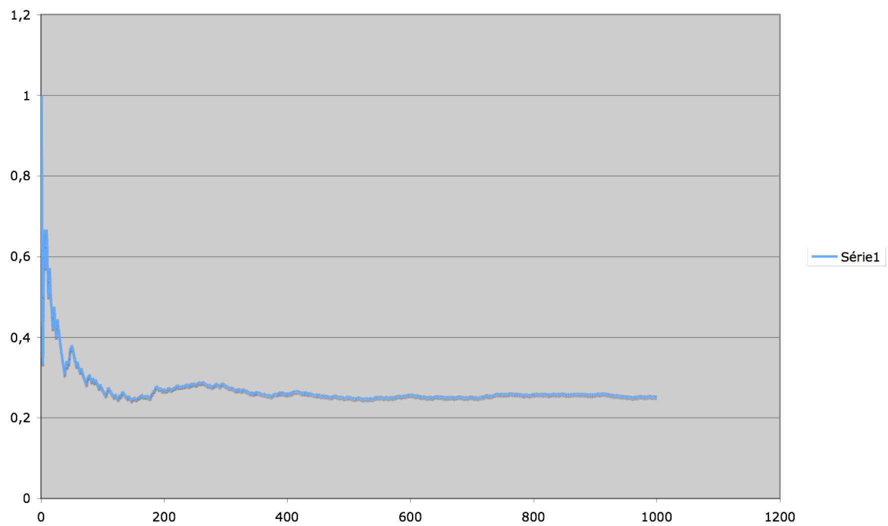
Programmes de première S, in B.O. n°7 Hors-série, 31 août 2000.

Nous avons déjà cité cet énoncé au paragraphe 2 de ce chapitre. On notera que les énoncés proposés dans les « programmes 2000 » de la loi des grands nombres évitent le terme de convergence pour la suite des fréquences, utilisant l'expression « se rapprochent de » ou « tendent vers », tendent entre guillemets.

Un mode conditionnel aurait été possible : « devraient se rapprocher de », marquant peut-être, comme nous l'avons évoqué au paragraphe 2 un choix dans l'interprétation du principe de Cournot, mais marquant peut-être aussi qu'il s'agit d'un autre type de convergence que la convergence « certaine ».

Par contre, ces mêmes programmes évoquent « *la convergence des moyennes vers l'espérance et des variances empiriques vers les variances théoriques ; on illustrera ceci par des simulations dans des cas simples* ». Il n'y a plus là de guillemets.

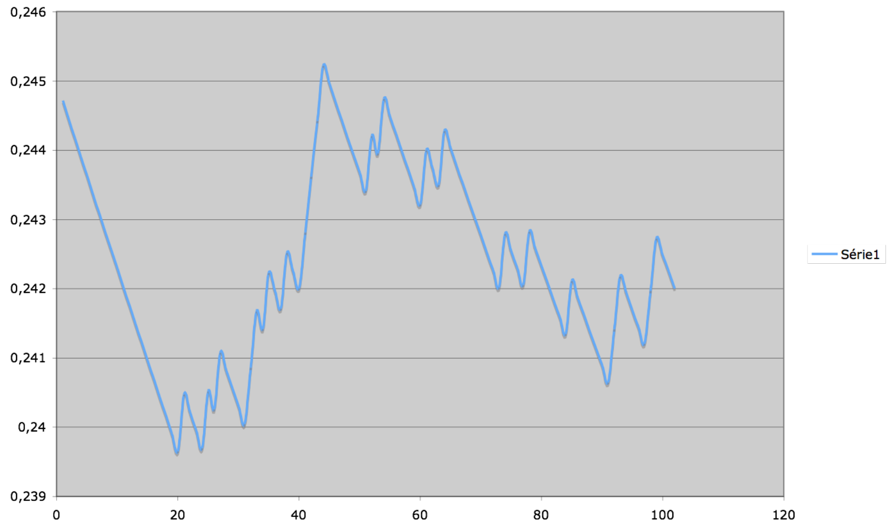
L'illustration de la loi des grands nombres se fait souvent, utilisant les facilités des tableurs, en présentant le nuage des points de coordonnées (n ; f_n) sur un graphique :



Ce graphique représente la suite (f_n) pour la simulation sur tableur de 1000 lancers d'un dé tétraédrique équilibré, f_n étant la fréquence des sorties observées de la face marquée 1 à l'issue des *n* premiers lancers. (En fait, on a ici une courbe reliant ces points, les points n'étant pas marqués dans un souci de lisibilité puisqu'ils sont très proches les uns des autres).

Il est vraisemblable que cette courbe présentée en dehors d'un cours de probabilité pourrait être interprétée comme représentant une suite (f_n) convergente au sens de l'analyse (convergence « certaine »).

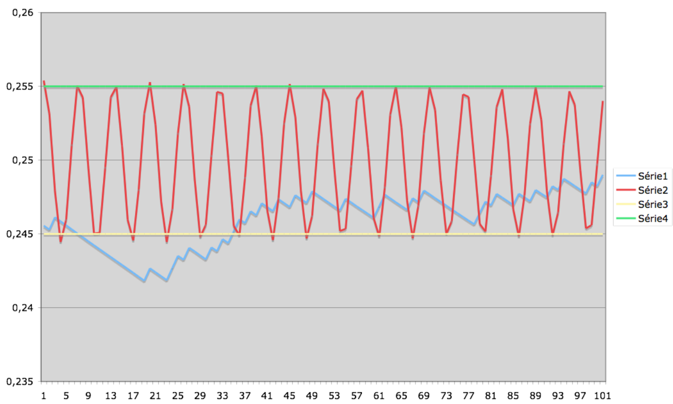
Ce type de représentation pour « convaincre » de la convergence joue grandement sur l'effet d'échelle (le choix de l'échelle sur l'axe des ordonnées étant géré par le tableur en fonction des valeurs extrêmes que prend la suite). Ainsi le graphique suivant représente pour **la même suite**, les 100 derniers termes calculés :



S'agit-il encore d'une suite « certainement » convergente ?

Ce phénomène peut s'observer d'ailleurs aussi pour la convergence « certaine ».

Dans le schéma ci-dessous, y a-t-il une suite convergente ? Oui, la suite représentée en rouge…



Il s'agit ici des termes de rang 900 à 1000 pour :

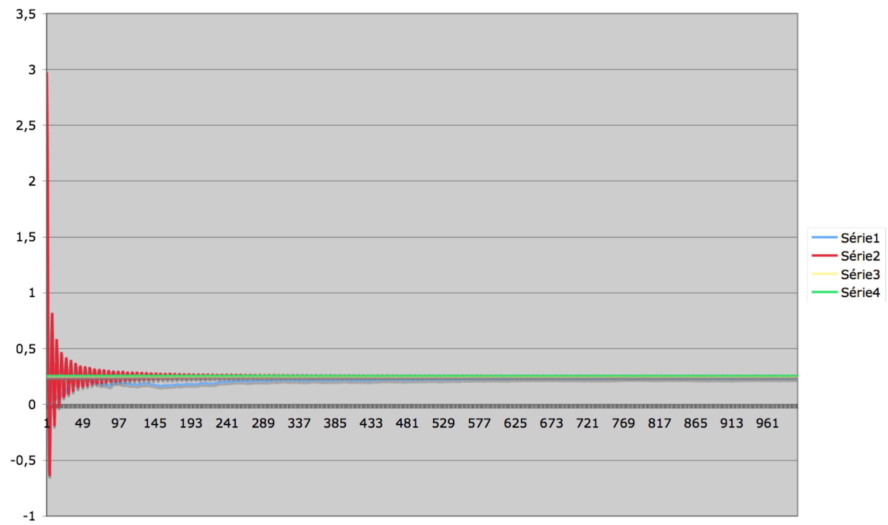
• La suite de terme général $u_n = (n+20*\sin(n))/(4*n+2)$, en rouge, suite qui converge vers 0,25

• La suite de terme général f_n égal à la fréquence de sortie de la face 1 lors de la simulation d'un dé tétraédrique équilibré, en bleu, qui converge en probabilité vers 0,25 également (c'est un autre échantillon).

• Les suites constantes égales à 0,245 et 0,255, en jaune et en vert.

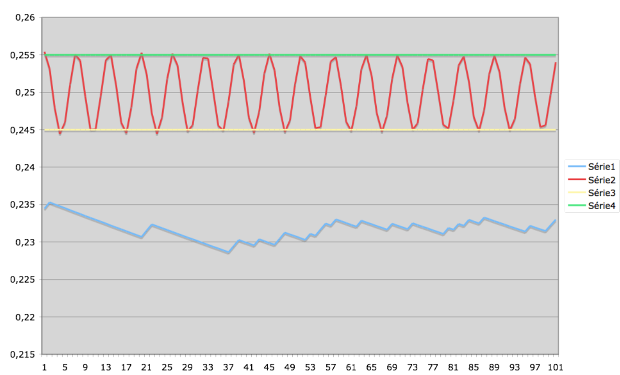
Les segments vert et jaune délimitent donc la bande définie par $0{,}245 \leq y \leq 0{,}255$: à partir d'un certain rang, tous les termes des deux suites (u_n) et (f_n) sont dans cette bande, enfin au moins jusqu'au rang 1000.

Le choix de l'échelle influe donc particulièrement sur l'interprétation qu'on fait « spontanément » du graphique. Or ce choix, l'acteur professeur ou l'acteur élève en est a priori dessaisi : la convergence certaine des mêmes quatre suites se lit sans aucun doute sur le graphe suivant représentant les 1000 premiers termes de chacune de ces quatre suites …

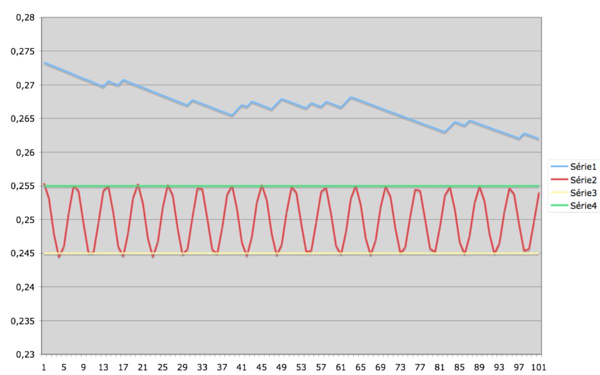


On peut, avec d'autres échantillons, obtenir pour les termes de rang 900 à 1000 :

ceci :



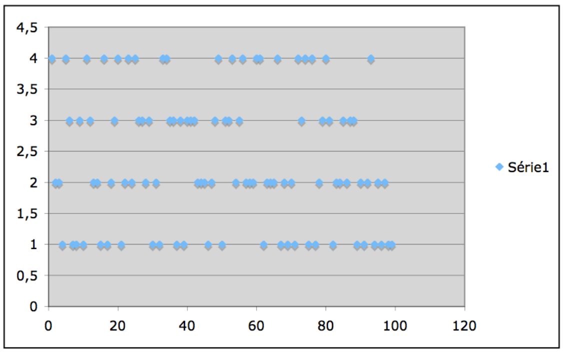
ou ceci :



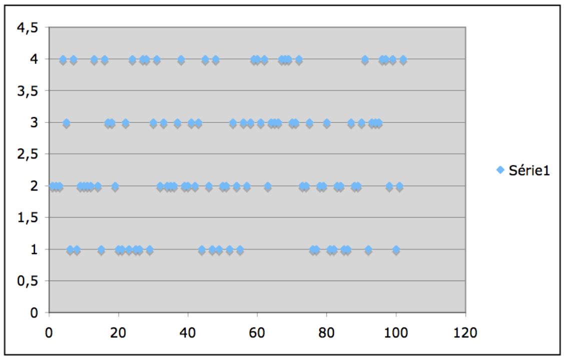
Ce qui illustre bien la fluctuation d'échantillonnage.

Par ailleurs, la représentation des issues pour les 100 premiers lancers est analogue à celle des 100 derniers lancers, la stabilisation de la fréquence sur le grand nombre de répétitions ne doit pas faire oublier que le désordre est le même, que ce soit sur les 100 premiers lancers, les 100 derniers, ou, si l'on itère l'expérience 10 000 fois, les 100 derniers de cette série de 10 000 :

pour les 100 premiers (1000 épreuves) :



et pour les 100 derniers même échantillon) :



L'ordonnée des points bleus (1, 2, 3 ou 4) représente le numéro de la face du dé sortie lors de l'essai dont le rang est en abscisse.
Il y a toujours autant de désordre.

L'illustration de la loi des grands nombres par les seules représentations graphiques de fréquences tendant à se stabiliser, si elle devait amener la confusion entre convergence certaine et convergence en probabilité, pourrait générer des conceptions erronées sur le hasard : finalement, pour un intervalle donné autour de p, on est sûr qu'à partir d'un certain rang f_n sera dans cet intervalle ; en quelque sorte, le hasard serait contrôlé, à condition d'être constant dans l'effort, en répétant un grand nombre de fois l'expérience.

L'outil informatique, au travers des simulations préconisées par les programmes, peut être utilisé de diverses manières pour illustrer la loi des grands nombres :

- par des représentations graphiques, avec les risques que nous venons d'évoquer,
- par des résultats numériques : calcul de la fréquence d'un événement A à partir de la simulation de n épreuves, où n est grand.

Un tel calcul effectué dans un langage impératif ne fera pas a priori apparaître de résultats intermédiaires [33], et gommera les fluctuations de la fréquence « au cours du temps », c'est à dire au fil des épreuves. Le tableur a l'avantage de permettre raisonnablement de visualiser la valeur de la fréquence obtenue à chaque nouvelle épreuve ; cependant lorsque le nombre d'épreuves répétées devient assez grand, une telle étude des valeurs numériques est assez fastidieuse, et la fluctuation s'amenuisant,en principe, au fil des épreuves, la convergence a tendance à nouveau à primer.

6. Le problème de la taille de l'échantillon

Comment choisir la taille de l'échantillon ? Plus grand est le nombre d'épreuves, meilleure devrait être l'approximation de la probabilité d'un événement par la fréquence observée de cet événement, mais même si l'ordinateur permet de réaliser un très grand nombre d'épreuves, ce nombre est fini.

Les problèmes à résoudre concernant une suite numérique (u_n) qui converge certainement vers un nombre p, et une suite de fréquence (f_n) qui converge en probabilité vers ce même nombre p, ne sont pas les mêmes :

• pour (u_n), un écart $e > 0$ étant donné, à partir de quel rang *n* est-on assuré que tous les u_n vont se trouver dans l'intervalle $]p - e ; p + e[$?

• pour (f_n), la taille *n* de l'échantillon étant donnée, pour quelle valeur de $e > 0$ est-on, sinon sûr, du moins presque sûr que p est dans $]f_n - e ; f_n + e[$?

La problématique concernant les suites certainement convergentes a toujours été introduite dans les programmes d'une manière ou d'une autre [34] ; la problématique pour la suite des fréquences commence à être introduite dans les « programmes 2010 » (voir le chapitre 3).

Nous observons dans nos classes de première et terminale S (programmes 2001 et 2002) un certain désarroi quant au choix de la taille de l'échantillon : « il faut répéter 100 fois, 1000 fois, plus ? ».

Le nombre de répétitions n'est pas une « bonne question ».

---

[33] sauf si on a prévu de les mémoriser.

[34] Y compris lorsque ne figurait pas au programme la définition formelle de la notion de suite convergente : beaucoup de sujets de baccalauréat approchaient un nombre singulier ß (zéro d'une dérivée ou autre) par la méthode du point fixe, et l'exercice se terminait par la question : comment choisir n pour que u_n approche ß à 10^{-3} près ?

La « bonne question » est : *n* étant fixé, quelle précision e peut-on espérer obtenir pour la probabilité p de l'événement A, en approchant p par la fréquence f_n d'apparition de A observée en répétant *n* fois l'expérience, et peut-on préciser un degré de confiance quant à l'appartenance de p à l'intervalle $]f_n - e ; f_n + e[$ ainsi déterminé ?

Si l'on a une relative liberté dans le choix de *n*, le plus grand possible en tenant compte de diverses contraintes : peut-on choisir *n* pour que, à un niveau de confiance donné, p soit dans l'intervalle $]f_n - e ; f_n + e[$?

Il n'y a pas de réponse à la question « combien de fois faut-il répéter l'expérience » sans définir au préalable la proximité visée, et le niveau de confiance attendu.

CHAPITRE 3

Les simulations dans les « programmes 2000 et 2010 »

La notion de simulation d'une expérience aléatoire apparaît dans ces programmes, pour la première fois dans les programmes de l'enseignement secondaire (mis à parts les programmes de terminale ES de 1994 à 1998, cf. chapitre 2, paragraphe 2) .

1. Les programmes des années 2000, 2001 et 2002

Les programmes de lycée ont changé en 2000 (seconde), 2001 (première) et 2002 (terminale). Ils ont été conçus par un groupe d'experts, dirigé par Claudine Schwartz, professeur des universités (Joseph-Fourier, Grenoble).
Ces experts sont essentiellement des universitaires et des professeurs de lycée .
Il est à noter que jusqu'alors les programmes étaient conçus sous l'autorité de l'Inspection Générale de Mathématiques, qui n'est intervenue pour ces nouveaux programmes que par l'un de ses représentants, membre du groupe d'experts.

Ces programmes se présentent sous la forme d'un document par niveau et série (plusieurs rubriques par document : analyse, géométrie, etc.) publié au Bulletin Officiel de l'Education Nationale ; chaque document est présenté dans des tableaux à trois colonnes : « contenus », « mise en œuvre » et « commentaires ».
Ces tableaux sont précédés de quelques pages destinées à les éclairer : « Généralités à propos d'une formation scientifique en première et terminale S », « Mathématiques et informatique en première et terminale S », « Epistémologie et histoire des mathématiques », « Organisation de l'enseignement et du travail des élèves » et « Les contenus du programme de première S », telles sont les rubriques de l'introduction aux programmes de première S.

Peu après la publication de ces programmes, des comités d'experts ont fourni des « documents d'accompagnements » (un pour chaque niveau : seconde, première et terminale, puis une « annexe » de probabilités / statistiques commune aux séries S et ES), documents assez conséquents (en comparaison avec les « commentaires des programmes » des années 1970-1980 regroupés parfois, toutes séries et tous niveaux confondus, en une seule brochure).

Ces documents d'accompagnement peuvent aussi bien donner un cadre théorique au professeur, que présenter des contextes, des exemples d'exercices, des prolongements au cours.

Pour ces « nouveaux » programmes a été affichée la volonté de concevoir des programmes pour les trois disciplines scientifiques enseignées en lycée (Sciences Physiques et Chimie, Sciences de la Vie et de la Terre et Mathématiques) qui soient non seulement cohérents entre eux, mais également permettent l'interdisciplinarité, et qui soient plus proches des pratiques des scientifiques du début du XXIème siècle.

Les trois groupes disciplinaires d'experts ont effectivement collaboré, et des thèmes sont communs à au moins deux disciplines scientifiques.

Par exemple :

• la désintégration nucléaire abordée d'abord en classe de physique, doit être la motivation pour l'introduction de l'équation différentielle $y' = y$, qui amène l'introduction de la fonction exponentielle en classe de mathématiques ;

• il doit être démontré en classe de mathématiques qu'une loi à durée de vie sans vieillissement est une loi exponentielle ; ce théorème peut alors être appliqué, en sciences physiques, à la désintégration des noyaux ;

• le lien doit être fait entre la datation au carbone 14 (programme de SVT) et la demi vie d'une loi exponentielle.

Mais surtout, là où, pour schématiser un peu, on lisait autrefois dans les programmes : « mise en équations » de problèmes concrets, puis « mathématisation » de problèmes, il est fait référence explicitement à une démarche de modélisation et, fréquemment, à des activités de simulation .

Le programme de statistiques de la classe de seconde est introduit par un préambule assez ambitieux :

> *« En seconde, le travail sera centré sur : la réflexion conduisant au choix de résumés numériques d'une série statistique ; la notion de fluctuation d'échantillonnage vue ici sous l'aspect élémentaire de la variabilité de la distribution des fréquences ; la simulation à l'aide du générateur aléatoire d'une calculatrice. La simulation remplaçant l'expérimentation permet, avec une grande économie de moyens, d'observer des résultats associés à la réalisation d'un très*

grand nombre d'expériences. (…)
En classe de Première et de Terminale, dans toutes les filières, on réfléchira sur la synthèse des données à l'aide du couple moyenne, écart-type qui sera vu à propos de phénomènes aléatoires gaussiens et par moyenne ou médiane et intervalle interquartile sinon. On amorcera une réflexion sur le problème de recueil des données et sur la notion de preuve statistique ; on fera un lien entre statistiques et probabilités L'enseignement de la statistique sera présent dans toutes les filières mais sous des formes diverses. »

Les programmes de seconde ont été publiés un an avant les programmes de première et terminale. Les ambitions concernant la réflexion sur le problème de recueil des données et la notion de preuve statistique ont été revues à la baisse lors de la publication des programmes de première et terminale.

Les programmes des années 2000 concernant les simulations, tels qu'ils sont définis dans les B.O. se résument facilement :
• En seconde est prescrite la mise en œuvre par les élèves de simulations de situations simples : lancer d'un dé (obligatoire) ou somme des points amenés en lançant deux dés (thème de statistique possible). Les outils utilisés peuvent être de type informatique (calculatrice ou tableur) ou des tables de nombres aléatoires. Ces simulations doivent permettre des calculs statistiques.
• Le programme de première S et ES prescrit des simulations légèrement plus complexes .
• Le programme de TS, TL et TES ne prescrit plus la mise en œuvre de simulations, mais l'utilisation de simulations (fournies) pour mettre en œuvre un test d'adéquation de la loi équirépartie à des données expérimentales.

Il est intéressant de voir quelle place ont les simulations dans les sujets du baccalauréat.
Ceci est important dans la mesure où beaucoup considèrent que ce sont les sujets du baccalauréat qui « pilotent » l'enseignement des mathématiques :
• Le troisième point (le test d'adéquation) a été évalué au baccalauréat en série S comme en série ES dès la session 2003, avec des exercices isomorphes entre eux : on présente un contexte « concret », qui fournit des données expérimentales sous forme d'une distribution des fréquences de réalisation des k issues d'une expérience aléatoire pour n répétitions de cette expérience. On fournit une série statistique de, par exemple, 1000 simulations de n

répétitions d'une loi équirépartie à k issue, ou certains indicateurs statistiques d'une telle série. Il est demandé de décider, au risque 5 % ou 10 %, si l'on accepte pour l'expérience aléatoire l'hypothèse d'une loi équirépartie. On trouvera en annexe de ce chapitre des énoncés de baccalauréat comportant la mise en œuvre de ce test.

• Lors de l'expérimentation limitée (année 2006/2007) d'une **« épreuve pratique »** qui devait compter pour un quart de la note à la session 2009[35] du baccalauréat S, l'un des sujets demandait à l'élève de faire une simulation sur tableur et une comparaison des résultats obtenus avec ceux qu'il avait obtenus par un calcul théorique :

sujet 011 — Épreuve pratique de mathématiques — Fiche élève

Simulation d'une expérience aléatoire, lois de probabilités

Énoncé

On dispose d'une roue divisée en trois secteurs identiques numérotés 1, 2 et 3.
On suppose qu'après rotation, la roue s'arrête sur l'un des trois secteurs de façon équiprobable.
On fait tourner successivement trois fois de suite la roue dans le sens trigonométrique en supposant que chaque résultat est indépendant des deux autres.
S désigne la variable aléatoire définie par la somme des trois numéros obtenus.
La variable aléatoire D est le numéro obtenu lors de la seconde rotation.

1. Sur un tableur réaliser une simulation de taille 100 de cette expérience.

   Appeler l'examinateur en cas de difficulté et pour valider.

2. Déterminer pour cette simulation les répartitions des fréquences de la variable aléatoire S.

   Appeler l'examinateur pour valider les résultats.

3. En utilisant les résultats connus sur la répétition d'expériences indépendantes, déterminer les lois de probabilités des variables aléatoires S et D.
4. La simulation du 2. est-elle cohérente avec les valeurs théoriques obtenues au 3. ?
5. Les évènements « S=3 » et « D=1 » sont-ils indépendants ?

Production demandée

– Pour les questions 3 et 5, les réponses sont à justifier.
– Pour la question 4, une rapide explication de la cohérence est demandée.

[35] Après avoir été repoussée à la session 2010, il est question d'attendre une refonte de l'examen pour la session 2013.

L'Inspection Générale, dans son compte-rendu de cette expérimentation n'a donné aucun commentaire sur ce sujet là (réponses attendues, erreurs rencontrées etc.) Il ressort cependant du compte-rendu que ce sujet a été parmi les sujets les moins souvent proposés par les enseignants à leurs élèves, et de loin. Les enseignants engagés dans l'expérimentation avaient une certaine liberté dans le choix, dans une « banque », imposée, de sujets à préparer avec leurs élèves.

L'expérimentation de cette « future » épreuve pratique qui fut menée au cours des années suivantes 2007/2008 et 2008/2009, a proposé plus de sujets comprenant une simulation d'expérience aléatoire (on trouvera certains de ces sujets en annexe de ce chapitre). Il n'y a pas eu, à notre connaissance, de compte-rendu sur ces expérimentations.

En conclusion, les raisons de cette apparition d'activités de simulations d'expériences aléatoires dans la prescription sont évoquées dans les B.O. et les documents d'accompagnement.

Nous citons ci-dessous pratiquement intégralement [36] le point 5 intitulé « Les contenus du programme de première S » du texte d' introduction aux programmes des séries scientifiques, *B.O. hors série n°7 du 31 août 2000, Volume 5, p. 29 et 30 de l'édition papier.*

> ***« Un programme se doit de répondre aux spécifications de formation fournies par l'institution, d'une part (représentée, en particulier, par le Conseil national des programmes) et, dans une certaine mesure et de façon peut-être moins explicite, par la communauté scientifique, d'autre part.***
>
> *Les spécifications notifiées par le Conseil national des programmes en janvier 1999 étaient d'introduire de la statistique en première et terminale S et d'utiliser les possibilités offertes par l'informatique. La prise en compte de cette demande, des attentes exprimées lors de la phase préparatoire à la rédaction de ce programme et, comme indiqué plus haut, la recherche d'un équilibre entre le poids des nouveautés, la continuité à assurer avec les anciens programmes et la faisabilité pour une classe d'âge donnée, ont conduit aux choix de contenus présentés dans les tableaux ci-après.*
>
> *(...)*
>
> *Désormais, la statistique est étudiée en série S ; aussi, quelques éléments sont-ils*

[36] Nous avons fait ressortir en caractères gras ce qui nous paraît particulièrement significatif.

> *développés sur cet enseignement (la longueur du commentaire n'est pas proportionnelle au temps à consacrer à ce sujet).*
> ***L'usage de la statistique dans de nombreux domaines ne relève pas d'une mode passagère mais de la diffusion d'une culture et d'un mode de pensée anciens, diffusion rendue possible par les progrès simultanés de la théorie mathématique et de la technologie informatique.*** *Chaque domaine d'application a une pratique spécifique de la statistique, fondée sur une problématique propre, le type d'expériences réalisables, la nature et les propriétés des données à traiter, les techniques de calcul mises en œuvre (on parle ainsi de statistique médicale, de statistique industrielle, de statistique financière, de physique statistique, etc.).*
> *En classe de seconde, les élèves ont acquis une expérience de l'aléatoire en pratiquant eux-mêmes des expériences de référence (lancers de dés, de pièces) et en simulant d'autres expériences à l'aide de listes de chiffres au hasard produites par une calculatrice ou un ordinateur.* ***La simulation joue un rôle important : en permettant d'observer des phénomènes variés, elle amène les élèves à enrichir considérablement leur expérience de l'aléatoire et favorise l'émergence d'un mode de pensée propre à la statistique ; elle rend de plus nécessaire la mise en place de fondements théoriques.***
> *Une partie du programme des classes de première et de terminale S concerne la modélisation d'expériences de référence, modélisation permettant d'expliquer des résultats observés ou d'en prévoir d'autres.* ***En première, on approfondira la notion de simulation d'une expérience, qui consiste à choisir un modèle et à le simuler ; la simulation permet, d'une part, d'avoir des estimations de résultats impossibles à calculer explicitement et, d'autre part, par la comparaison de telles estimations avec des résultats expérimentaux, de valider le modèle choisi.***
> *La statistique descriptive a une part modeste dans la série S ; en particulier, on n'aborde pas les problèmes de recueil des données qui varient considérablement d'un domaine à l'autre ; ces questions font l'objet d'enseignements spécifiques dans les études qu'un élève de S est susceptible d'entreprendre ultérieurement. »*

On remarquera que sur les 34 lignes de cette citation, mises à part les 8 lignes (tronquées ici) consacrées à la description des tableaux qui définissent le programme, 26 lignes sont consacrées à justifier la place de la statistique et l'introduction de simulations.

Les éléments mis en avant sont, pour ce qui concerne la place de la statistique, d'ordre sociaux : la nécessaire prise en compte de l'évolution des sciences et des techniques (« *les possibilités offertes par l'informatique* »), et la multiplicité des champs d'application.

La place des simulations est justifiée de manière plus confuse.
Les simulations doivent démarrer en classe de seconde (alors que l'enseignement des probabilités ne commence officiellement qu'en classe de première) : il s'agit de travailler la notion de variabilité de la fréquence (ou la fluctuation d'échantillonnage), notion plutôt rattachée à la statistique certes, mais qui met nécessairement en jeu de l'aléatoire lors de la constitution d'un échantillon de taille n sur lequel calculer des indicateurs statistiques. Ce qui est proposé ici, c'est d'aller directement vers des échantillons de n épreuves d'une même expérience aléatoire : lancer d'une pièce, d'un dé.
Il est donc proposé implicitement de travailler, non pas sur une population « palpable » dont on veut étudier certains caractères (taille, couleur des yeux) grâce à une étude statistique à partir d'échantillons de taille n choisis aléatoirement, mais sur des données aléatoires qui ne préexistent pas avant l'étude, créées au fur et à mesure en répétant l'expérience n fois.

C'est un choix qui permet d'éviter le difficile problème de la constitution de l'échantillon, mais qui a l'inconvénient, en ramenant l'enseignement de la statistique à l'étude statistique de résultats d'expériences aléatoires, de ne peut-être pas favoriser autant que le souhaitaient les concepteurs des programmes « *l'émergence d'un mode de pensée propre à la statistique* » pour reprendre une formulation que l'on trouve dans les programmes, et que nous développerons dans la quatrième partie.
Evidemment dans ce cadre là, il est plus commode de travailler sur des simulations que sur des expériences réellement observées : faire simuler 1000 lancers d'un dé équilibré avec un tableur prend moins de temps et fait moins de bruit que faire lancer 1000 fois un dé par 35 élèves.

La simulation est considérée ici, **implicitement encore**, comme elle-même une expérience aléatoire puisque « *en permettant d'observer des phénomènes variés, elle* [la simulation] *amène les élèves à enrichir considérablement leur expérience de l'aléatoire* ».

Le morceau de phrase « *elle* [la simulation] *rend de plus nécessaire la mise en place de fondements théoriques* » est plutôt sibyllin.

Est-ce une manière de suggérer que les élèves ne pourront véritablement « intégrer » le principe des simulations qu'après des travaux théoriques à venir ?
On doit en effet remarquer que, à part ce bout de phrase plutôt ambigu, les liens à établir avec les élèves entre une réalité de caractère aléatoire et une simulation de cette réalité ne sont pas même évoqués. Rappelons qu'il s'agit de la fin du paragraphe décrivant les activités de la classe de seconde, alors que l'enseignement des probabilités ne commence qu'en première.

Ou est-ce une manière de dire qu'il s'agit de placer l'élève de seconde en situation d'être demandeur d'un « fondement théorique » à ces simulations, au travers de l'enseignement des probabilités de la classe de première ?
Le premier paragraphe de l'annexe 1 du document d'accompagnement des programmes, classe de première des séries générales (p. 67) pourrait confirmer cette seconde interprétation :

« *Les élèves ont ainsi observé des résultats qui appellent une explication. De telles explications relèvent des mathématiques, mais nécessitent de formaliser et de préciser dans le cadre d'une théorie (celle des probabilités) le langage utilisé pour parler de l'aléatoire* »

Le dernier paragraphe de la citation du B.O. propose une définition de la notion de simulation d'une expérience : « *[elle] consiste à choisir un modèle et à le simuler* » ; autrement dit, simuler une expérience aléatoire, c'est simuler son modèle.

La définition est circulaire.

Cependant, il y a un renversement dans l'utilisation de l'expression *simulation*, un glissement de "*simulation d'une expérience aléatoire*", qui renvoie à une certaine réalité, à : "*simulation d'une loi de probabilité*", qui renvoie à la théorie ; il arrive qu'on rencontre une phrase telle que "le lancer d'un dé cubique équilibré est une simulation de la loi uniforme sur {1 ; 2 ; ... ; 6}" (cf. chapitre 1, paragraphe 2).
Cette formulation paraîtrait assez pertinente dans ce contexte : la loi uniforme sur {1 ; 2 ; ... ; 6} est une loi idéale, virtuelle, inaccessible dans la réalité. Si l'on veut s'en approcher, créer une réalité qui ressemble à cette loi, une réalité qui la simule, on peut lancer un dé, <u>qu'on va supposer *équilibré*</u>. L'expérience est, elle, effective, et l'hypothèse « *équilibré* » faite sur le dé s'entend : c'est un dé non pipé, qui a été fabriqué pour être équilibré.

Nous émettons l'hypothèse que dans l'esprit des concepteurs des programmes, la deuxième occurrence du verbe simuler, « simuler le modèle » peut s'entendre par exemple ainsi : le modèle peut être rendu isomorphe au modèle d' un tirage de boules d'une urne de composition bien choisie, et c'est cette expérience aléatoire de tirage de boules d'une urne qui est la *simulation du modèle* ; sa réalisation effective peut être « effectuée à la main » ou remplacée par un programme informatique [37], en ayant mis en place un dispositif expérimental qui rende plausible l'hypothèse du tirage *au hasard* (urne opaque, boules « indiscernables au toucher », brassage des boules dans l'urne, nombres pseudo-aléatoires « de qualité », etc.). La mise en scène du « tirage du loto » est censée réaliser ce genre de conditions expérimentales.

Dans le cas d'une loi non équirépartie, le programme se limite à

> *«* [la simulation de lois] *de probabilités simples obtenues comme images d'une loi équirépartie par une variable aléatoire (sondage, somme des faces de deux dés, etc.). »*

(colonne centrale « modalité de mise en œuvre » du tableau définissant le contenu du paragraphe « probabilité et statistique » du programme de Première S, B.O. référencé plus haut).

> *« On indiquera que simuler une expérience consiste à simuler un modèle de cette expérience. La modélisation avec des lois ne découlant pas d'une loi équirépartie est hors programme »*

(idem, colonne de droite « commentaires)

La définition de la simulation est toujours résolument circulaire, au moins dans sa forme.

Le B.O. précise également l'utilisation qui peut être faîte des simulations :

> *« la simulation permet d'une part d'avoir des estimations de résultats impossibles à calculer explicitement et d'autre part, par la comparaison de telles estimations*

[37] Nous emploierons souvent cette expression « programme informatique » pour désigner un procédé attaché à un outil de calcul qui peut aussi bien être une calculatrice de lycée, qu'un tableur, l'utilisation d'un logiciel dédié ou la programmation dans un langage procédural d'un algorithme : nous n'entrerons pas, dans cette première partie de notre travail, dans des analyses sur les tâches et procédures différentes nécessitées par tel ou tel choix d'outil.

avec des résultats expérimentaux, de valider le modèle choisi. »

La deuxième partie de cette dernière citation renvoie au programme de terminale S, plus précisément au test d'adéquation du modèle équiréparti aux données expérimentales qui y est proposé : ce test consiste à comparer, grâce à un indicateur convenablement choisi, les données expérimentale liées à l'expérience aléatoire initiale, avec les données expérimentales obtenues par une simulation de la loi de probabilité équirépartie. Nous aurons l'occasion d'analyser dans la quatrième partie ce test.

La première partie de la citation paraît plus problématique, car peut-être contradictoire avec la définition qui vient d'être proposée d'une simulation : si simuler une expérience aléatoire consiste à simuler un modèle de cette expérience, comment avoir des estimations de résultats grâce à la simulation, alors qu'on est dans une situation où les résultats sont impossibles à calculer explicitement, parce que l'on n'arrive pas à déterminer une loi de probabilité de l'expérience ?

On peut cependant envisager de trouver un modèle par tâtonnements : on choisit un modèle (les procédures pour déterminer un tel modèle restant à préciser), on le simule et on confronte les résultats observés lors de simulations, avec les résultats expérimentaux de l'expérience d'origine. Cette confrontation pourrait permettre de valider le modèle, ou de l'invalider : il faudrait alors chercher un autre modèle.

Les sujets de probabilité proposés à l'épreuve pratique expérimentale précédemment évoquée contredisent la définition proposée d'une simulation : la simulation de l'expérience aléatoire y précède toujours la détermination de sa loi de probabilité, sous le titre « étude expérimentale ».

En conclusion, la lettre des programmes, telle qu'elle est publiée au Bulletin Officiel, et telle qu'elle s'incarne dans les sujets de baccalauréat, élaborés sous la responsabilité de l'Inspection de Mathématiques (autorité garante de la mise en œuvre de ces programmes), ne définit pas clairement ce qu'est une simulation d'une expérience aléatoire dans le cadre de l'enseignement général des lycées.
Nous allons essayer d'analyser ce que disent des simulations les documents d'accompagnement de ces programmes.

2. Les simulations dans les documents d'accompagnement (« programmes 2000 »)

Nous allons citer des extraits des documents d'accompagnement de seconde, premières et terminales où sont évoquées les simulations et les modèles probabilistes d'une expérience aléatoire.

Nous citons de longs passages de ces documents, ce qui va donner à ce paragraphe un caractère de paraphrase, que nous assumons.

Cela paraît nécessaire pour faire apparaître la cohérence de ces « programmes 2000 » concernant les simulations d'expériences aléatoires, cohérence qui s'articule autour de cette définition de la simulation :

> *« Formellement, simuler une expérience, c'est choisir un modèle pour cette expérience – ici une loi de probabilité P sur un certain ensemble – puis produire une liste de données pour lesquelles P est un modèle pertinent »*

(Programme de Seconde publié dans le B. O. hors-série, n°6, du 12 août 1999).

2.1. Utilisation des simulations

Nous avons fait ressortir certains passages en caractères gras.

> *« On peut observer par simulation de p séries de taille n d'une même expérience que la variance s_g de la série des p moyennes est plus petite que la variance des np données (ou plus simplement situer s_g dans la série des p variances des séries) : la moyenne d'une série est une quantité variable, mais «moins variable» que les données elles-mêmes »*

(annexe 1 du document d'accompagnement de Première S)

La simulation d'expériences aléatoires apparaît ici comme un procédé commode pour construire des échantillons sur lesquelles effectuer des calculs statistiques (afin de mettre en évidence un résultat de statistique), en contournant la difficulté de la constitution d'échantillons d'une population, difficulté déjà évoquée.

> *« L'exemple ci-dessous montre comment mêler diverses composantes d'un travail mathématique: observation, premières conjectures, expérimentation à plus grande échelle, puis obtention et preuve de certains résultats.*
>
> *Exemple : somme de deux dés*

En répétant cent fois de suite le lancer de deux dés et en effectuant la somme des points obtenus, on observe que certains résultats s'obtiennent plus souvent que d'autres.
À l'aide d'un tableur, par exemple, il est possible d'expérimenter à plus grande échelle : simulation d'un plus grand nombre de lancers de deux dés et construction du tableau des effectifs . L'inégale répartition des fréquences de chaque résultat est flagrante.
***L'algorithme de simulation du lancer de deux dés sera conçu par l'élève** et d'éventuelles erreurs (simuler un lancer de dé et doubler le résultat, simuler une loi uniforme sur {2,..., 12}) seront analysées.*
La recherche d'un modèle théorique permet ensuite de faire les calculs. »

Résultats pour 5 séries de 100 lancers

2	0,02	0,02	0,01	0,03	0,01
3	0,03	0,02	0,09	0,06	0,04
4	0,08	0,08	0,08	0,08	0,1
5	0,11	0,11	0,12	0,09	0,11
6	0,17	0,13	0,1	0,11	0,11
7	0,15	0,15	0,12	0,14	0,2
8	0,12	0,2	0,1	0,13	0,22
9	0,15	0,1	0,12	0,13	0,06
10	0,09	0,1	0,14	0,1	0,08
11	0,06	0,05	0,06	0,09	0,06
12	0,02	0,04	0,06	0,04	0,01

Le document d'accompagnement spécifie bien que l'activité de simulation qui vient d'être décrite, est menée par l'élève : « ***l'algorithme de simulation du lancer de deux dés sera conçu par l'élève*** ».
Remarquons que c'est l'un des rares moments des divers documents accompagnant ou commentant les programmes où il est fait mention d'une activité de type algorithmique à propos des simulations.
Les procédures mises en jeu par l'élève pour créer cet algorithme ne sont pas développées.

C'est l'élève qui doit concevoir la simulation : cela peut permettre d'agir sur le biais d'équiprobabilité ; un algorithme faux, tel que l'un des deux algorithmes suggérés dans la citation ci-dessus, manifestera que l'élève ne semble pas avoir dépassé l'obstacle de l'équiprobabilité ; peut-être la confrontation entre les résultats obtenus en lançant pour de bon les dés, et les résultats obtenus par la simulation, pourra-t-elle amener l'élève à se poser les bonnes questions.

Mais rien ne prouve que l'élève qui simule correctement le lancer des deux dés par la commande sur tableur « = Alea.entre.bornes(1 ; 6) + Alea.entre.bornes(1 ; 6) », a dépassé cet obstacle.

2.2. La définition de la simulation dans les documents d'accompagnement

La définition du B.O. est confirmée et précisée par les documents d'accompagnement :

> *« **Simulation de chiffres au hasard***
> *On clarifiera brièvement les positions respectives de la modélisation et de la simulation : modéliser consiste à associer un modèle à des données expérimentales, alors que simuler consiste à produire des données à partir d'un modèle prédéfini. On parlera de simulation d'une loi de probabilité P ; la simulation d'une telle loi avec des listes de chiffres au hasard ne peut se faire que si P peut être construite à partir d'une loi équirépartie. Pour simuler une expérience, on associe d'abord un modèle à l'expérience en cours, puis on simule la loi du modèle; on pourra détailler ces étapes, sans cependant le faire systématiquement dans les cas simples des expériences de référence. »*

(accompagnement Première S, à partir de la page 70)

3. Les « programmes 2010 »

Les programmes décrits aux paragraphes 1 et 2. de ce chapitre seront caducs à la rentrée scolaire 2010 pour la classe de seconde, à la rentrée 2011 [38] pour les classes de première et à la rentrée 2012 pour ceux des classes de terminale.

Les programmes de seconde ont été publiés au Bulletin Officiel n° 30 du 23 juillet 2009. Nous produisons en annexe de ce chapitre la partie concernant l'enseignement de la statistique et des probabilités.

---

[38] En fait, les programmes de seconde ont été changés dès l'années scolaire 2009-2010. Ceux de première et terminale des années scolaires suivantes sont adaptés à titre provisoire, pour prendre en compte ces changements, en particulier l'introduction des probabilités dès la classe de troisième pour cette classe d'âge, et une initiation à l'algorithmique programmation en classe de seconde.

Dans ces programmes, l'élève apparaît clairement comme l'acteur de la simulation.

L'objectif annoncé des simulations d'expériences aléatoires est, outre la mise en œuvre d'activités de type algorithmique, de créer des échantillons pour

> *« amener les élèves à un questionnement lors des activités suivantes :*
> *• l'estimation d'une proportion inconnue à partir d'un échantillon ;*
> *• la prise de décision à partir d'un échantillon. »*

Concernant l'estimation, le programme précise, en note :

> *« L'intervalle de fluctuation au seuil de 95%, relatif aux échantillons de taille n, est l'intervalle centré autour de p, proportion du caractère dans la population, où se situe, avec une probabilité égale à 0,95 , la fréquence observée dans un échantillon de taille n. Cet intervalle peut être obtenu, de façon approchée, par simulation. Le professeur peut indiquer aux élèves le résultat suivant, utilisable dans la pratique pour des échantillons de taille* $n \geq 25$ *et des proportions p du caractère comprises entre 0, 2 et 0, 8 : si f désigne la fréquence du caractère dans l'échantillon, f appartient à l'intervalle* $[p - 1/\sqrt{n}, p + 1/\sqrt{n}]$ *avec une probabilité d'au moins 0, 95. Le professeur peut faire percevoir expérimentalement la validité de cette propriété mais elle n'est pas exigible. »*

Les programmes de mathématiques de seconde font déjà une part importante aux simulations informatiques.

Notons que le lien, proposé par le document « Ressources pour la classe de seconde, Probabilités et Statistiques » (document rérérencé « [3] » à la fin de la bibliographie de cette thèse), entre modèle et simulation d'une expérience aléatoire, est stable par rapport aux programmes antérieurs : la modélisation précède la simulation.

Nous citons ci-dessous un extrait de ce document « Ressources pour la classe de seconde, », page 6 :

2.3. Modélisation, modélisations ?

La simulation a pour préalable de choisir un modèle.

Exemple 1 : somme de deux dés
Cette situation se prête volontiers à la mise en œuvre d'une démarche consistant à proposer un modèle et à le confronter aux données d'expérience. Les résultats compris entre 2 et 12 peuvent conduire certains élèves à faire porter l'équiprobabilité sur l'ensemble des 11 résultats observables. Par quelques expérimentations avec des dés, puis en ayant recours à des simulations, on est conduit à rejeter ce modèle pour proposer de faire porter l'équiprobabilité sur les 36 couples de résultats {(1,1) ;(1,2) …(6,6)}, présentés usuellement sous forme de tableau croisé.

Les simulations proposées dans l'exemple 1 sont des simulations de la loi (erronée) équirépartie sur {2 ; 3 ; … ; 12}, qu'on rejette après l'avoir confrontée avec « quelques expérimentations avec des dés » ; il s'agit donc bien de simulations du modèle choisi, et non de simulations de l'expérience aléatoire telle qu'on peut en lire dans la suite du document :
en effet, page 13 est présenté un algorithme (dans l'encadré bleu) de simulation d'une expérience aléatoire :

Exemple : nombre de lancers pour sortir tous les numéros d'un dé cubique[20]

On souhaite estimer le nombre de lancers nécessaires pour sortir toutes les faces d'un dé cubique.

Comme pour l'exemple précédent, la réponse est aléatoire. On peut, dans un premier temps, écrire un algorithme permettant de simuler une expérience, puis faire tourner plusieurs fois cet algorithme et observer que la moyenne des résultats obtenus se stabilise[21] pour un grand nombre d'expériences. Cette moyenne, elle aussi aléatoire, permet néanmoins de proposer une réponse au problème en termes de probabilités.

```
Algorithme
        Variables
              L liste
              S, n, x entiers
        Initialisations
              Vider la liste L
              S et n prennent la valeur 0
        Traitement
              Tant que[22] n est inférieur à 6
              S prend la valeur S +1
              x prend la valeur d'un nombre aléatoire entier entre 1 et 6
              si x ne figure pas dans la liste L alors
                     n prend la valeur n+1
                     L(n) prend la valeur x
              Fin du Tant que
        Sorties
             Afficher S
```

On peut, pour estimer le nombre moyen de lancers permettant de sortir les six faces du dé, modifier l'algorithme précédent en s'inspirant de l'algorithme 2 ci-avant.

L'expérience aléatoire envisagée consiste à lancer le dé cubique tant que toutes les faces ne sont pas sortie au moins une fois. Il est proposé ici un algorithme de simulation de cette expérience, que l'on peut facilement traduire en un programme informatique dont l'exécution sera une simulation de l'expérience aléatoire. L'algorithme n'est d'évidence pas construit à partir d'un modèle probabiliste de l'expérience aléatoire simulée.

CHAPITRE 4

A quoi servent les simulations ?

1. Les simulations peuvent-elles valider le modèle probabiliste ?

Delahaye (2005), dans un article publié dans « Pour la Science » (n° 336, octobre 2005, p. 90 à 94) écrit en introduction de cet article :

> *« Les solutions paradoxales des plus simples problèmes de probabilité engendrent de passionnantes controverses que les simulations permettent de trancher. »*

Delahaye évoque dans le corps de son article des situations aléatoires pour lesquelles les calculs de probabilité donnent des résultats qui peuvent surprendre (« *solutions paradoxales* »). Il affirme que quelques simulations de l'expérience suffisent à convaincre de la pertinence du résultat (« *Jouez au jeu une dizaine de fois* »)

Le nombre réduit de simulations qui, selon l'auteur, permettent de convaincre le lecteur, nous a amenée à nous interroger sur l'intérêt des simulations, particulièrement lorsqu'on est capable de déterminer la loi de probabilité de l'expérience aléatoire que l'on veut simuler.

Les situations évoquées par Delahaye sont des expériences aléatoires pour lesquelles il a su construire un modèle probabiliste. Les simulations de ces expériences permettent de récupérer des données empiriques, et ces données empiriques illustreront les résultats du calcul mené dans le modèle, et par là, affirme l'auteur, convaincront les réticents.

Si, comme l'énoncent les programmes, la simulation est construite à partir du modèle probabiliste de l'expérience aléatoire, l'observation des résultats de la simulation de la situation initiale ne devrait pas dire plus que l'observation des résultats du calcul mené dans le modèle.
L'observation des résultats d'une dizaine de simulations risque même de dire moins que l'observation des résultats du calcul, du fait de la fluctuation d'échantillonnage, comme nous allons le voir à propos du « problème du Vizir » évoqué par Delahaye :

Le vizir propose à Sindbad d'ouvrir l'une des trois portes A, B et C. Il précise que derrière une seule des portes, on trouvera un trésor (dont l'emplacement a été choisi au hasard). Sindbad gagnera le trésor s'il ouvre finalement la bonne porte. Par exemple, Sindbad choisit la porte A, mais n'ouvre pas immédiatement cette porte. Le vizir, qui sait où est le trésor, ouvre une autre porte (B ou C) derrière laquelle il n'y a rien (il choisit la porte au hasard si le trésor n'est ni derrière B, ni derrière C). Sindbad peut alors décider de choisir d'ouvrir une autre porte que la porte A qu'il avait initialement choisie.
A-t-il intérêt à modifier son choix initial ?

Delahaye montre par un calcul ce qu'il présente comme un paradoxe : en changeant son choix initial, Sindbad double ses chances de gagner. Il écrit alors :

« Il existe heureusement une méthode ultime pour se persuader de la justesse de l'idée qu'il faut changer de porte : la simulation. Jouez au jeu une dizaine de fois en adoptant la stratégie « ne pas changer » et ensuite une dizaine de fois en adoptant la stratégie « changer » et faîtes le bilan. En cherchant sur Internet à partir des mots Monty Hall et Applet ***vous trouverez des programmes informatiques qui faciliteront ces simulations et qui normalement devraient vous convaincre qu'on double ses chances en changeant de porte ».***

Nous avons fait cette simulation, en programmant une calculatrice de lycée (TI 82). Il ne s'agit d'ailleurs pas réellement d'une simulation mais plutôt d'une implantation du jeu sur une calculatrice :

un nombre entier X entre 1 et 3 est tiré au hasard grâce à la fonction rand de la calculatrice ; on doit trouver ce nombre et on a droit à deux essais, la première proposition étant toujours 1 (le choix initial de Sindbad).
Si X est égal à 1, la calculatrice tire au hasard l'un des deux nombres 2 ou 3 et l'affiche.
On fait alors une deuxième proposition Y :
• si on change le premier choix, on propose celui des deux nombres 2 ou 3 qui n'a pas été affiché
• si on maintient le choix, on propose 1
Si Y=X, la calculatrice affiche « GAGNE », sinon elle affiche « PERDU ».

Pour quinze parties, en changeant le choix initial, « on » a gagné 11 parties sur 15 ; en maintenant le choix, on a gagné 2 parties sur 15.

Qui ne connaît rien au probabilités, conclura qu'on semble avoir plus de chance de gagner en changeant le choix qu'en le maintenant, et peut-être même conclura qu'en changeant de choix, on multiplie les chances par 5,5.
Qui s'y connaît un peu en statistique-probabilités, dira que l'échantillon est peut-être trop petit pour pouvoir conclure quoi que ce soit du fait de la fluctuation d'échantillonnage, et transformera le programme pour voir ce qui se passe pour un plus grand nombre de parties.

Contrairement à ce qu'affirme Delahaye, en jouant une dizaine de fois avec l'une et l'autre des stratégies, ces simulations , « normalement », ne devraient convaincre personne **qu'on double ses chances** en changeant le choix.
Nous sommes en train de travailler avec le hasard, et sur un échantillon d'effectif 15, la fluctuation d'échantillonnage s'illustre pleinement.

Delahaye décrit, dans ce même article, une autre situation qu'il appelle « le problème des Sophie », situation qu'il qualifie également de paradoxale. Il propose également, pour convaincre le lecteur resté sceptique devant le résultat du calcul, d'observer les résultats de quelques simulations.

2. Qu'est-ce qu'une simulation ? Une première approche.

Les simulations proposées dans cet article de Delahaye à propos du « problème des Sophie » nous ont paru non adaptées à l'expérience en jeu (nous ne présentons ici ni le problème, ni les simulations proposées).
Nous avons alors été alors amenée à nous interroger sur les critères qui peuvent permettre de considérer qu'une « simulation » est acceptable, et partant nous nous sommes interrogée sur la définition même d'une simulation d'une expérience aléatoire.

La définition proposée dans les programmes présente, comme nous l'avons souligné, plusieurs inconvénients :

• elle est circulaire,

• elle est contredite par la pratique :

- que ce soit la pratique révélée par l'exemple du paragraphe 3 du chapitre précédent,
- que ce soit la pratique imposée par les énoncés des épreuves pratiques expérimentales, où la simulation précède la détermination de la loi de probabilité.

Nous allons bousculer un petit peu cette définition selon laquelle la simulation d'une expérience aléatoire serait la simulation de la loi de probabilité qu'on lui associe.
La simulation d'une expérience aléatoire est nécessairement liée au modèle probabiliste de cette expérience : peut-être pourrait-on dire, en première approche, que la simulation d'une expérience aléatoire est une autre expérience aléatoire <u>que l'on peut réaliser</u> (et pour laquelle on pourra observer des résultats), et à qui on peut associer le même modèle probabiliste que pour l'expérience initiale.
Nous pourrons ainsi dire que tirer *au hasard* un jeton d'une urne contenant six jetons numérotés de 1 à 6 est une simulation du lancer d'un dé *équilibré*, mais aussi que tirer deux fois de suite *au hasard* un jeton de cette urne, en remettant le premier jeton tiré avant de tirer le second est une expérience aléatoire qui simule le lancer de deux dés *équilibrés*. Les expressions *équilibré* et *au hasard* renvoient ici à ce que nous avons déjà nommé au chapitre précédent un « protocole expérimental » : le dé n'est pas pipé, les jetons ont été brassés, l'acteur de l'expérience ne triche pas, etc.

L'expérience et sa simulation ont même modèle probabiliste, **que ce modèle ait été construit** (comme le suggèrent les « programmes 2000 »), **ou qu'il ne l'ait pas été**.
Mais nous butons là sur une difficulté théorique : comment peut-on dire que deux expériences aléatoires ont même modèle probabiliste si l'on n'a pas déterminé ce modèle ?

C'est en partie cette question qu'aborde Parzysz (2009) dans un article publié dans la revue REPERES-IREM, article que nous allons partiellement présenter et commenter.

3. L'approche de Parzysz (2009) :
« De l'expérience aléatoire au modèle, via la simulation »

Dans cet article Parzysz tente de définir les liens entre l'expérience aléatoire, le modèle que l'on en donne, et les simulations que l'on peut faire de l'expérience.
Son objectif au départ, qui justifie le titre de son article, est de faire prendre conscience aux

élèves qu'un certain nombre d'expériences aléatoires ont même modèle. Mais il s'agit justement d'introduire la notion de modèle probabiliste, donc le travail à mener doit éviter la référence à ce modèle probabiliste : il faut donc faire ressortir les similitudes de ces expériences, pour que l'élève puisse les considérer comme « équivalentes », sans rapporter cette équivalence au modèle probabiliste commun.

3.1. Un premier exemple

Parzysz propose d'associer à chacune de ces diverses expériences une simulation à l'aide d'un tableur, puis de faire constater aux élèves l'isomorphie des feuilles de calcul associées à chacune des expériences.

Parzysz suggère que l'on peut en préalable créer une situation didactique dans la classe pour introduire la nécessité de simuler une expérience (il prend l'exemple du jeu de Croix ou Pile : Parzysz (2007)) , pour répéter un grand nombre de fois l'expérience dans un délai raisonnable.

Mais se pose la question : qu'est-ce que simuler ?

Pour Parzysz, il s'agit de remplacer l'expérience aléatoire initiale par une autre expérience aléatoire, « *plus facile et/ou rapide à mettre en œuvre* », et « *qui reflète statistiquement l'expérience initiale, l'adéquation entre les deux étant en quelque sorte assurée par la qualité du modèle probabiliste déterminant la simulation* » ; il s'appuie sur une citation de Dodge :

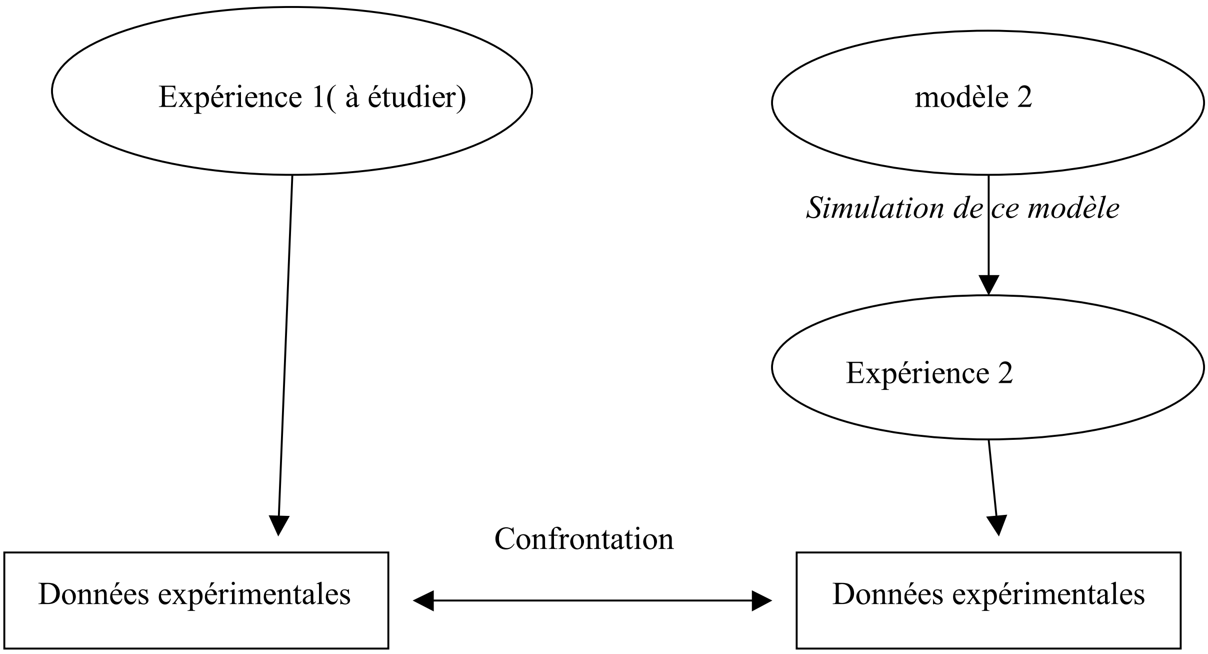
> *« La simulation est la méthode statistique permettant la reconstitution fictive de l'évolution d'un phénomène. C'est une expérimentation qui suppose la construction d'un modèle théorique présentant une similitude de propriétés ou de relations avec le phénomène faisant l'objet de l'étude ».* Dodge (1993)

Mais dans la situation didactique qu'il présente, le modèle théorique en question ne peut, pour une expérience aléatoire, être son modèle probabiliste, alors que l'auteur veut justement introduire cette notion.

Parzysz reprend à son compte la définition d'une simulation, ou plutôt de l'action de simuler, donnée dans le document d'accompagnement de seconde : « *simuler une expérience, c'est choisir un modèle de cette expérience puis simuler ce modèle »,* et propose, en classe de seconde de « court-circuiter » ce choix en faisant travailler directement les élèves sur la simulation [39].

[39] Parzysz ne précise pas qui, in fine, construit la feuille de calcul, c'est à dire la simulation.

Nous avons donc le schéma suivant :



La confrontation des données expérimentales obtenues avec les deux expériences permet d'accepter ou non l'expérience 2 comme simulation de l'expérience 1.

Expérience 2 est une expérience aléatoire qui a, par construction, pour modèle probabiliste : modèle 2.

Quel lien y a-t-il entre Expérience 1 et modèle 2 ? Parzysz reprend les termes de Henry (1999), il s'agit de définir un « *modèle pseudo-concret* » :

> *« une sorte d'abstraction et de simplification de la réalité perçue dans sa complexité, dans la mesure où certains choix sont faits pour ne retenir que ce qui semble pertinent de cette situation vis à vis du problème étudié »*

Et Parzysz illustre le choix de ce modèle pseudo-concret pour la simulation d'une série de lancers d'une pièce de monnaie, qui consiste à préciser ses hypothèses :

H1. la pièce peut tomber sur Pile ou sur Face ;

H2. si elle tombe sur la tranche, on annule le lancer [40] ;

H3. Pile et Face ont autant de chances d'être obtenus ;

H4. « On lance la pièce suffisamment fort pour que le résultat ne soit pas prévisible ».

L'ordinateur fournit un nombre aléatoire x selon la loi uniforme sur [0 ;1] ; Parzysz traduit ses

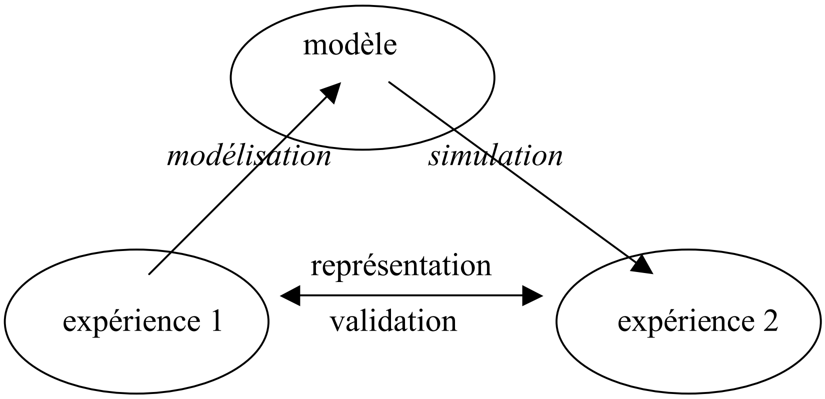
---

[40] Nous aurions plutôt proposé : « la pièce ne peut pas tomber sur la tranche », en combinant peut-être H2 et H4.Il s'agit là d'une illustration du principe fort de Cournot…Nous pensons qu'annuler un lancer pose des problèmes à certains élèves, comme les lancers virtuels évoqués plus loin par Parzysz.

hypothèses : H1 et H2 en associant Pile à l'événement x < 0,5 et Face à l'événement x ≥ 0,5 ; H3 est satisfaite puisque la loi de x est uniforme, et H4 aussi car on a confiance dans le générateur de nombres aléatoires.

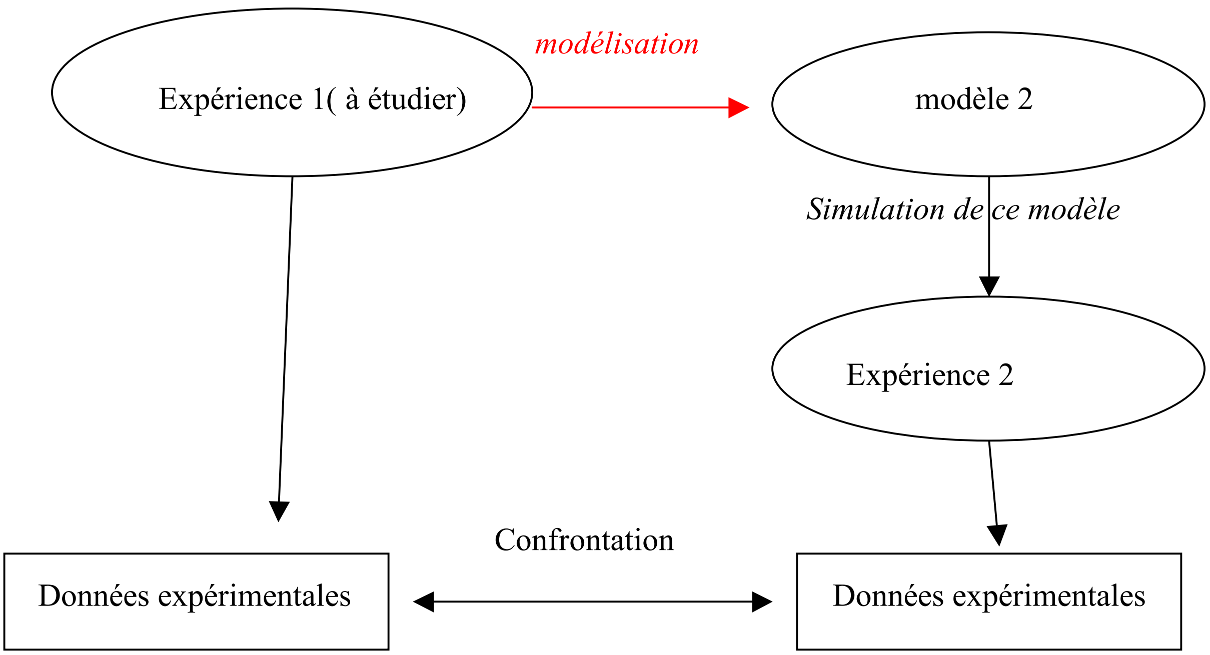
<u>3.2. Le schéma ternaire de Parzyscz</u>

Parzysz propose alors le « *schéma ternaire* » :



Autrement dit, dans le schéma que nous avions établi à la page précédente, il rajoute la flèche *modélisation*. (voir la figure ci-dessous).

L'activité de modélisation a consisté, dans le cas d'une suite de lancers de pièce, à émettre les hypothèses évoquées plus haut : *les hypothèses de modèle*.



Ce dernier lien *modélisation* est présenté de façon assez vague concernant la classe de seconde : les hypothèses de modèle sont simultanément traduites et interprétées par l'activation du générateur aléatoire.

On peut remarquer que pour l'expérience très élémentaire du lancer d'une pièce, les hypothèses de modèle sont, ni plus ni moins un modèle, le modèle 2 du schéma, qui est aussi un modèle pour l'expérience initiale (le lancer d'une pièce) : loi équirépartie pour les issues du lancer.
Ces hypothèses de modèle semblent tout à fait accessibles à des élèves de seconde pour ce qui concerne H1, H2 et H3.

L'interprétation de H4 en terme d'indépendance des essais est plus problématique : sont en jeux justement la notion difficile d'indépendances d'événements (Maury (2005, 2006), Nabbout (2006)) et les biais évoqués au chapitre 2.
Cette hypothèse d'indépendance, c'est le « spécialiste » qui l'a rajoutée, spécialiste qui doit par ailleurs composer avec le fait que la machine ne génère que des nombres pseudo-aléatoires et qui se demande si les résultats de la fonction qui génère l'alea, peuvent être considérés comme indépendants.
Cet aspect n'a peut-être pas à être abordé avec les élèves, en tout cas pas avec des élèves débutants, qui justement sont susceptibles avoir des connaissances spontanées erronées sur l'indépendance des épreuves lorsque l'on lance une pièce sous l'hypothèse H4 (par exemple le biais de la loi des petits nombres).
D'autant que cette hypothèse H4 est posée, non pour la simulation à proprement parler du lancer de la pièce, mais pour la répétition de cette simulation. La question de l'indépendance des essais ne se pose pas pour les élèves au moment de la construction de la simulation *stricto sensu* du lancer de la pièce, et ne devrait pas se poser quand il s'agira de répéter la simulation, c'est à dire de recopier les formules vers le bas dans la feuille de calcul [41].

Donc la modélisation, sous sa forme H1, H2 et H3, est une tâche qu'un élève de seconde peut mener à bien, sans savoir qu'il a déterminé la loi de probabilité d'une expérience aléatoire.

Parzysz écrit :

« *Le problème est qu'il faut attendre la classe de Première pour aborder*

[41] Le hasard est, **dans ce contexte**, assez « désincarné » : nos observations spontanées nous amènent à faire l'hypothèse que les élèves n'essayent pas de se représenter le processus qui amène ces nombres « aléatoires ». Les élèves ne nous ont jamais posé de question sur ce problème d'indépendance, **dans ce contexte**.

véritablement la notion de modèle, et la question de savoir si une simulation est acceptable demeure entière à ce stade », [la classe de seconde],

et le seul lien possible entre l'expérience et sa simulation, en l'absence d'une référence possible au modèle, est de confronter les résultats expérimentaux que donnent l'une et l'autre.

La procédure de confrontation n'est pas décrite par Parzysz : nous supposons que cette confrontation passe par la répétition de l'expérience.

Parzysz n'aborde pas la question de la taille des échantillons à considérer, ici, échantillon pour l'expérience initiale, et échantillon pour la simulation, non plus que la question du ou des critères qui permettront « *d'établir si une simulation fournit bien les même résultats que l'expérience* » en tenant compte des fluctuations d'échantillonnage.

La question de la validation de la simulation paraît donc rester, dans ce cadre non résolue : que doit-on faire observer aux élèves pour qu'ils établissent, à partir des données expérimentales, *un lien entre l'expérience et une simulation de cette expérience* comme dit le document d'accompagnement des programmes ?

Pour conclure sur le « schéma ternaire » proposé par Parzyscz, nous comprenons que le *modèle* dont il est question est un modèle de l'expérience aléatoire initiale.

Parzysz avait écrit, avant de proposer son schéma, commentant la définition de Dodge :

> *« Ceci ne veut cependant pas dire que pour simuler une expérience on doit connaître son modèle probabiliste (alors il n'y aurait plus de problème, sauf de calcul). La démarche est en fait la suivante : vu le protocole expérimental et en fonction des outils que j'ai à ma disposition, je propose un modèle pour cette expérience. Les données expérimentales, confrontées aux simulations de ce modèle, me permettent ensuite (ou non) d'accepter ce modèle comme adéquat ou de le rejeter »*

Au départ, donc, on n'a pas de modèle probabiliste (sinon, simuler ne pose pas de problème, serait même peut-être inutile). En tenant compte du protocole expérimental, et « en fonction d'outils à disposition », on construit un modèle de l'expérience (qu'on simule et qu'on valide, ou non, en confrontant les données expérimentales de l'expérience et de la simulation du modèle construit).

La méthode qui permet de construire dans ces conditions un modèle de l'expérience que l'on veut étudier n'est pas précisée.

Il demeure donc deux questions :

• que fait-on en l'absence de données expérimentales sur l'expérience initiale ? Comment alors valider la simulation et partant le modèle qu'elle simule ?

• que signifie finalement « simuler un modèle » dans le cas où l'on n'a pas accès « spontanément » à son modèle probabiliste. Comment construit-on ce modèle qu'on simule ?

3.3. Un autre exemple

L'auteur présente quatre expériences :

E1 : 4 lancers successifs d'une pièce et on note le nombre de « Pile » obtenus.

E2 : On lance 4 pièces simultanément et on note le nombre de « Pile » obtenus.

E3 : Un sac contient 5 billes blanches et 5 billes noires. On effectue 4 tirages successifs d'une bille (avec remise) et on note le nombre de billes blanches obtenues.

E4 : On laisse tomber une bille dans une planche de Galton à 5 godets et on note le numéro du godet dans lequel elle tombe

Parzysz écrit :

> *« Pour simuler chacune de ces quatre expériences, nous allons nous efforcer de déterminer une procédure* ***aussi "proche" que possible de celle de l'expérience elle-même*** [42]*, et nous la mettrons en œuvre sur le tableur .»*

Pour chacune des quatre expériences, Parzysz définit ses *hypothèses de modèle*, puis donne quatre procédures, c'est à dire comme nous l'avons souligné plus haut : quatre algorithmes, qui le mènent cette fois à **quatre tableaux identiques**, correspondant à des feuilles de calcul identiques, aux notations près.

Nous ne présenterons pas plus avant cet essai.

Parzyscz a présenté et illustré une méthode pour construire en classe de seconde des simulations d'expériences aléatoires en restant dans le cadre imposé par les programmes (la définition qui est donnée des simulations, et l'introduction des lois de probabilité seulement en classe de première), en ne cachant pas les difficultés pour garder une cohérence.

---

[42] souligné par nous : nous aurons à revenir sur cette idée de « procédure aussi proche que possible de l'expérience ». La mise entre guillemets de *proche* par Parzyscz souligne la difficulté à définir cette idée de proximité.

Par ailleurs il suggère une méthode, de façon vague certes, qui ouvre de nouvelles pistes : déterminer pour simuler une expérience aléatoire une procédure « aussi proche que possible de celle de l'expérience elle-même ».

Dans le résumé de son article, il parlait de « *la mise en oeuvre de procédures de simulation qui soient* ***aussi « isomorphes » que possible*** [43] *à l'expérience simulée et l'explicitation des hypothèses qui sous-tendent ces procédures.* »

Il ne s'agit plus alors de construire la simulation à partir du modèle probabiliste de l'expérience aléatoire à simuler.

[43] souligné par nous.

CHAPITRE 5

Des travaux de chercheurs s'appuyant sur des simulations

Nous allons présenter brièvement trois thèses qui sont adossées à des simulations *informatiques* d'expériences aléatoires.
Nous présenterons d'abord la thèse de Zaki (1990) qui souhaite offrir à ses étudiants une possibilité d'expérimentation, grâce à des simulations, pour accompagner des résolutions de problèmes ; nous présenterons ensuite quelques éléments de la thèse de Bordier (1991) qui a imaginé un « laboratoire de probabilité », fondé sur des expérimentations à l'aide de simulations informatiques.
Nous présenterons enfin la thèse de Coutinho (2001) qui conçoit une ingénierie didactique pour l'enseignement des probabilités fondée sur des simulations.

1. La thèse de Zaki (1990)

Le sujet en est : « traitements de problèmes de probabilités en situation de simulation ».
Zaki part du constat que dans l'enseignement des probabilités par les définitions axiomatiques, ses étudiants rencontrent des difficultés, d'une part à associer cette définition à leurs conceptions initiales des probabilités et d'autre part à donner une interprétation aux éléments du modèle (univers, tribu, probabilité) et à leur relations. Il écrit :

> *« Devant tous ces problèmes, nous avons opté pour des observations d'assez longue durée auprès d'étudiants de première année scientifique d'Université, portant sur des situations probabilistes tributaires de paramètres (données numériques) et donnant lieu à deux entrées :*
>
> - *La résolution de quelques cas particuliers simples, préalablement à l'étude des situations dans des cas plus complexes,*
> - *La simulation, à l'aide de l'ordinateur, des situations pour différentes valeurs des données numériques, en vue de faire des estimations. »*

Il remarque que les situations de prévision et de décision présentées aux élèves l'ont été jusqu'alors surtout par des schémas d'urnes ou des équivalents. Ces situations sont assez éloignées des situations couramment rencontrées (jeux, sondages etc.), et éludent l'étape de la modélisation probabiliste, « *qui est l'étape fondamentale et spécifique des probabilités* » : il s'agit de trouver un modèle probabiliste qui rende compte « *parfaitement* » de la situation étudiée. Cette étape précède ce que Zaki appelle *le traitement déductif* qui permet de résoudre le problème posé.

> *« C'est cet ordre contraignant des tâches, modélisation puis traitement déductif, qui confère à l'enseignement des probabilités sa particularité par rapport au reste des mathématiques. Nous pensons qu'un enseignement des probabilités ne peut pas atteindre son objectif sans une bonne initiation à la modélisation probabiliste. Et cette initiation suppose que l'on ne se limite pas à l'approche a priori des probabilités, mais que l'on aborde le point de vue "fréquentiste" ; c'est ce point de vue qui s'impose par exemple pour donner des estimations avant le dépouillement total du scrutin d'une opération électorale. »*

Zaki a placé ses étudiants dans des situations concrètes, qui n'étaient pas des situations de décision ou de prévision, et qui nécessitaient un

> « *traitement combinatoire complet (éventuellement appliqué à des espaces d'épreuves infinis), et dans lesquelles les notions d'indépendance stochastique ou de probabilités (conditionnelles ou non) prennent un caractère tout à fait naturel* ».

Pour étudier ces situations, Zaki a mis à la disposition des étudiants des simulations informatiques de tirages d'urnes.

Les situations choisies par Zaki peuvent être modélisées par un schéma d'urnes (une ou deux urnes), où le tirage se fait avec ou sans remise.
Ce sont ces tirages qui sont simulés, l'étudiant ayant la faculté de définir la composition de l'urne ou des urnes (pour chaque exercice, c'est le logiciel qui indique le nombre d'urnes) ; l'étudiant peut choisir le nombre de répétitions des tirages, et le type d'exploitation statistique qu'il souhaite obtenir sur les résultats ainsi générés.
Zaki écrit :

> « [La démarche de simulation] *permet de :*
>
> - *"voir" l'évolution du phénomène aléatoire auquel se réfère la situation probabiliste retenue ;*
> - *faire des estimations (probabilités, espérances, intervalles de confiances) ;*
> - *contrôler des résultats mathématiques, des représentations et des*

modélisations préétablis (sans que cela soit un moyen de résolution). »

Cela n'est possible que s'il y a répétition de la simulation :

> « *On peut considérer qu'il y a là une barrière, qui doit être franchie pour que la simulation serve. Cela nous conduit à une seconde hypothèse : la simulation va permettre aux étudiants qui sont au delà de cette barrière d'enrichir leur conception probabiliste. Néanmoins, cet enrichissement ne va pas être uniforme, mais lié aux questions que les étudiants vont se poser en faisant la simulation* » Zaki [1990]

Les étudiants observés par Zaki ont travaillé en binôme.
Les binômes qui ont répété les simulations (cela n'a pas été le cas de tous les binômes), les ont utilisées pour faire des estimations de probabilités par une approche fréquentiste. Certains d'entre eux ont noté des écarts entre les résultats donnés par le calcul dans le modèle probabiliste et les résultats de ces estimations ; ils ont attribué ces écarts au peu de simulations réalisées, et ont indiqué qu'ils auraient eu de meilleures estimations en faisant plus de simulations. Pour Zaki, ces étudiants ont une idée intuitive « *qui va dans le sens des définitions statistiques de la probabilité* » [44].
Zaki estime que la pratique de la simulation « *a eu quelques effets favorables sur le développement du jugement probabiliste des étudiants* » et pense que l'expérimentation décrite dans sa thèse confirme qu'il est possible d'utiliser cette pratique des simulations pour identifier certaines conceptions probabilistes erronées.

2. La thèse de Bordier (1991)

Bordier (1991) constate, après d'autres chercheurs, que beaucoup d'élèves ou d'adultes ont des idées préconçues fausses concernant le hasard, et liste ces erreurs.
Certains élèves peuvent même avoir des conceptions théoriques correctes, mais qu'ils contredisent dans leurs pratiques.
Bordier conteste l'efficacité des remèdes proposés jusqu'alors : un enseignement théorique des probabilités ou un enseignement fondé sur la mise en œuvre par l'élève d'expériences aléatoires ou un enseignement mixte.

---

[44] Rappelons que pour ces étudiants, l'entrée dans l'enseignement des probabilités était une entrée par une présentation axiomatique.

Selon lui, la mise en œuvre par l'élève d'expériences aléatoires permettant de corriger les conceptions fausses bute sur la taille nécessairement insuffisante à son avis de l'échantillon que l'élève peut constituer.

Bordier propose donc de créer un « laboratoire de probabilité », où un certain nombre d'expériences aléatoires qu'il juge pertinentes en regard des conceptions à corriger, peuvent être simulées. Ce laboratoire repose sur un logiciel informatique fermé [45], et permet d'avoir rapidement, et sous des formes variées, les résultats d'un grand nombre de simulations de chacune de ces expériences.

Son laboratoire fonctionne suivant le modèle théorique suivant :

Réalité (situation aléatoire) → modélisation (en termes de tirages de boules dans des urnes) → simulation à l'aide du didacticiel → confrontation des résultats de la simulation aux conceptions initiales pour éventuellement les déstabiliser.

La modélisation telle que Bordier l'envisage, crée non pas un modèle (au sens de la théorie des probabilités : un espace probabilisé), mais une nouvelle situation aléatoire, équivalente à la précédente, en termes de tirages d'une ou plusieurs boules, avec ou sans remise, d'une ou plusieurs urnes.

Nous aurions tendance à dire que cette nouvelle situation aléatoire obtenue au terme de la phase de modélisation est elle-même une simulation de la réalité initiale, ce qu'exprime implicitement Bordier (p. 47) en posant dans un ensemble de tests, la question suivante, pour laquelle il a choisi comme titre : « <u>La perception de l'équivalence des contextes</u> ».

> *« Pour utiliser le laboratoire de probabilité, décrit au chapitre 4, les élèves doivent être en mesure de représenter les problèmes à résoudre dans un contexte de tirage de boules dans des bouliers[46] ; Or il n'était pas évident pour nous que de tels changements de contexte seraient perçus comme allant de soi par les élèves. En conséquence, il était important, avant que ne commence la réalisation du logiciel, de juger de l'aptitude des élèves à réaliser de telles transpositions.*

---

[45] Au sens où l'élève peut faire un choix limité de paramètres dans des situations de tirages de boules d'une urne. Ce logiciel est dédié à la simulation de situations aléatoires à des fins pédagogiques. On a parlé un temps de didacticiel.

[46] « boulier » : urne en « canadien »

• Vous avez décidé de jouer, avec un groupe d'amis, une partie de Monopoly ; or vous ne trouvez pas les dés. Pouvez-vous trouver un moyen équivalent de jouer au Monopoly ?

Pour une majorité d'élèves (80 %) le problème est apparu comme très simple ; on remplaçait, généralement, les lancers des dés par deux tirages, avec remise, d'un billet parmi 6 étiquetés de 1 à 6 ; d'autres utilisaient deux chapeaux avec 6 billets dans chacun. Notons qu'aucun élève n'a proposé de tirer deux billets d'un coup d'un seul chapeau. Les réponses, où aucun substitut aux dés ne fut indiqué, ne permettent pas d'affirmer que ces élèves étaient incapables d'une telle transposition. En effet, la majorité de ces derniers répondirent : « je jouerais à un autre jeu » ou encore " je n'aime pas le jeu de monopoly " »

Sans l'assurance que ce changement de cadre se fait « sans problème », l'efficacité du laboratoire peut être compromise.

Une des démarches de Bordier est assez originale, et à notre connaissance peu utilisée en France : pour une expérience aléatoire donnée présentée de manière pas nécessairement mathématisée, il définit un événement A lié à cette expérience et pose la question : « Combien de fois faut-il <u>en moyenne</u> répéter l'expérience pour que l'événement A se produise ? »
L'élève doit alors passer par cette étape que Bordier qualifie de « modélisation » pour simuler l'expérience (en termes de tirages d'urnes), et préciser le nombre de répétitions de l'expérience (« tant que A n'est pas survenu ») à l'aide des primitives du logiciel. L'élève peut demander autant de simulations qu'il veut ; le didacticiel se charge d'établir toutes les statistiques que l'élève demandera sur ces simulations.

Si l'événement A a pour probabilité p, si on appelle X le nombre de fois où il a fallu répéter l'expérience avant que A ne se produise pour la première fois, on démontre aisément que l'espérance de X est $E(X) = 1/p$.
En effet, X prend un nombre infini de valeurs : 1, 2, …., la valeur k est prise avec la probabilité $(1-p)^{k-1}p$: on suppose bien sûr que les essais sont deux à deux indépendants ; si A ne survient qu'au $k^{ème}$ essai, c'est que lors des k-1 premiers essais, on a eu non(A), avec la probabilité $1 – p$, A survenant au $k^{ème}$ essai.
$E(X) = p \sum k (1-p)^{k-1}$, somme infinie qui se calcule aisément en dérivant la série

$f(x) = \sum x^k = \frac{1}{1 - x}$; on a : $f'(x) = \sum k\, x^{k-1} = \frac{1}{(1 - x)^2}$,

d'où $E(X) = \frac{1}{p}$.

Il est clair que les élèves ne sauraient mener ce calcul.
Avec le laboratoire de probabilité, les élèves voient la valeur moyenne du nombre d'essais nécessaires à la survenue de A se stabiliser vers un nombre dont Bordier pense pouvoir faire découvrir aux élèves qu'il est l'inverse du nombre que dans certaines situations ils ont pu associer à l'événement A, par diverses considérations.

On peut remarquer ici que la thèse de Bordier est essentiellement théorique quant aux séquences qu'il imagine avec des élèves, puisqu'il n'a expérimenté son laboratoire, et son modèle d'apprentissage qu'avec quatre élèves (niveau seconde), et encore seulement avec une partie de son laboratoire.

Bordier propose « un modèle didactique pour l'enseignement de la probabilité » (chapitre 5) qu'il résume par « *trois étapes bien distinctes :*
• *L'identification des conceptions intuitives des élèves ;*
• *La construction par l'élève, au moyen d'outils appropriés, d'une science expérimentale de la probabilité ;*
• *La formalisation de la «théorie expérimentale.* »

Bordier pense que les simulations sans ordinateur avec des tables de nombres aléatoires, présentent des inconvénients majeurs : durée de la mise en œuvre et de son traitement, et valeurs obtenues assez loin des valeurs attendues.

Il envisage cependant une difficulté avec les simulations informatiques, due au fait que l'ordinateur ne produit que des nombres pseudo-aléatoires :

> « *D'une part, il existe de bons générateurs aléatoires et, d'autre part, des objets concrets tels des pièces de monnaie, malgré leur imparfaite symétrie, exhiberont peu de biais, si elles sont convenablement lancées. Reste que pour un élève, le hasard produit par l'ordinateur ne sera pas nécessairement perçu comme équivalent à celui des expériences réelles avec des objets ; ainsi avant d'utiliser*

un logiciel de simulation, il faudra faire constater aux élèves, sur plusieurs exemples très simples, que le hasard produit par le programme est équivalent à celui des expériences correspondantes . »

3. La thèse de Coutinho (2001)

Dans sa thèse « *introduction aux situations aléatoires dès le collège : de la modélisation à la simulation d'expériences de Bernoulli dans l'environnement informatique Cabri-géomètre II* », Coutinho postule que pour des adolescents (15-16 ans), l'acquisition de la notion de modèle probabiliste (sous la forme d'une loi de probabilité sur les issues) doit être précédée d'une étape dans un *domaine pseudo-concret* (Henry (2001)), où un *modèle d'urne* est associé à l'expérience aléatoire étudiée.

Définir ce modèle d'urne consiste à indiquer la composition de l'urne, et le type de tirage à y effectuer (avec ou sans remise). En fait, Coutinho se limite à des expériences du type *expériences de Bernoulli*, c'est à dire des expériences à deux issues : succès avec la probabilité p et échec avec la probabilité $q = 1 - p$. ; l'urne à associer à une telle expérience est caractérisée par la proportion p, dans l'urne, des boules blanches qui représentent le succès.

Ce modèle d'urne doit permettre de définir une notion nouvelle pour les élèves : la *pré-probabilité*, notion fonctionnant dans le domaine pseudo-concret des urnes, et qui est, pour les élèves, la proportion p des boules dans l'urne des boules blanches caractérisant le succès.
Bien sûr cette proportion p est, dans le modèle probabiliste, la probabilité du succès.

Définir le modèle d'urne adéquat à associer à une expérience aléatoire revient donc à définir une expérience aléatoire « équivalente » à l'expérience initiale, expérience consistant à tirer une boule d'une urne de composition convenable.

L'ingénierie didactique imaginée par Coutinho a donc pour objectif d'amener des élèves non encore au stade du concept de probabilité, à déterminer dans trois situations du type « expérience de Bernouilli », la proportion p de boules blanches à mettre dans l'urne pour que le tirage d'une boule de l'urne soit une expérience équivalente à l'expérience étudiée ; l'équivalence entre expériences n'est pas encore institutionnalisée pour ces élèves [47] :

« *La séance a commencé par un bref rappel sur les notions d'expérience*

[47] ni dans les situations didactiques qui précèdent l'exercice, ni dans les consignes de l'exercice.

aléatoire, d'urne de Bernoulli et d'expériences équivalentes (que l'enseignant a présentées comme des expériences comparables) » Coutinho (2001) p. 209.

Et c'est bien la reconnaissance de l'équivalence de deux expériences aléatoires que vise Coutinho.
Elle attend des élèves qu'ils fassent la différence entre *hasard* et *contingence*, et pour cela essaye de faire attacher au terme d'*expérience aléatoire* deux caractéristiques : d'une part le caractère de *reproductibilité*, avec intervention du *hasard*, et d'autre part la description de deux issues possibles (rappelons qu'elle limite son ingénierie à des expériences de Bernoulli).
Elle écrit p. 221 :

> *« En conséquence, l'identification de deux expériences équivalentes n'est pas encore un savoir-faire opératoire chez ces élèves car il leur manque les connaissances qui constituent les critères de validité pour une telle opération :*
> *a) Ce sont des expériences aléatoires*
> *b) Elles sont représentées par la même urne de Bernoulli (les résultats possibles sont organisés selon la même proportion). »*

Pendant l'expérimentation de l'ingénierie didactique, l'élève est censé rester dans le domaine pseudo-concret et y déterminer des pré-probabilité, qu'il peut cependant exprimer en termes de « chances », par une fraction ou en pourcentage.

L'élève peut essayer de définir la proportion p (pré-probabilité) par un raisonnement souvent d'ordre géométrique ; il peut aussi s'appuyer sur des simulations de la situation pré-programmées sur ordinateur : la simulation est répétée N fois. L'élève peut ainsi comparer la fréquence de succès lors de ces N essais avec la proportion p qu'il avait déterminée par le raisonnement, ou décider de sa proposition pour p à partir de l'observation des résultats des simulations.
Coutinho reconnaît que ceci suppose une utilisation de la loi des grands nombres, également non institutionnalisée dans son ingénierie.

Chacun des trois exercices principaux constituant l'ingénierie didactique décrite, peut-être plutôt en fait trois étapes dans une progression pédagogique, est préparé à l'aide de *situations didactiques* (Coutinho se réfère à Brousseau) qui amènent les notions nécessaires à la bonne marche de l'étape suivante. Coutinho présente également ces trois étapes comme des situations didactiques.

Coutinho tient à mener de front les deux approches, laplacienne et fréquentiste. Elle n'explique pas nettement dans la description de son ingénierie comment les élèves font ce lien, en particulier comment leur vient l'idée de s'appuyer sur des simulations répétées des expériences aléatoires étudiées pour déterminer la composition de l'urne.

L'expression *simulation de l'expérience* est employée dans les explications ou les consignes données aux élèves, mais là encore sans définition institutionnelle.

Dans son chapitre III, Coutinho écrit :

> *« le dispositif "urne à pixels" est l'outil de simulation de l'expérience de référence "tirage au hasard d'une perle dans un pot rempli de perles rouges et bleues" »*,

Ce dispositif, qui « discrétise » le plan, permet d'amener les élèves, pour l'expérience aléatoire dérivée du jeu de « franc carreau » à calculer la « pré-probabilité » à associer à l'expérience comme quotient de deux aires.

4. Quelques remarques sur ces travaux

Ces trois thèses ont un certain nombre de points en commun :

• Elle mettent en jeu des simulations informatiques de tirages d'urnes, tirages d'urnes qui sont eux-mêmes des simulations des expériences aléatoires à étudier : les expériences aléatoires d'origine doivent d'abord être « modélisées » en termes de tirages d'urnes.

• Elles s'appuient sur des simulations informatiques d'expériences aléatoires produites par des logiciels fermés : ce ne sont pas les élèves qui ont conçu ces simulations ; l'action des élèves est limitée aux compositions des urnes, au choix du nombre de répétitions et des outils statistiques pour exploiter les résultats obtenus.

• Elles décrivent chacune une forme d'ingénierie didactique, s'appuyant sur des simulations informatiques d'expériences aléatoires.

En revanche, les objectifs assignés à ces simulations ne sont pas du tout de même nature, même si les trois auteurs pensent qu'il est indispensable d'associer une démarche expérimentale à l'enseignement des probabilités.

• Zaki veut offrir une démarche expérimentale (visant en particulier une approche fréquentiste) à des étudiants qui reçoivent un enseignement essentiellement axiomatique de la notion de probabilité d'un événement.

• Bordier veut utiliser les simulations informatiques pour corriger certaines conceptions erronées des élèves : de telles simulations peuvent être répétées un grand nombre de fois, ce qui n'est pas le cas avec des simulations non informatiques, trop lentes ou compliquées à mettre en œuvre.

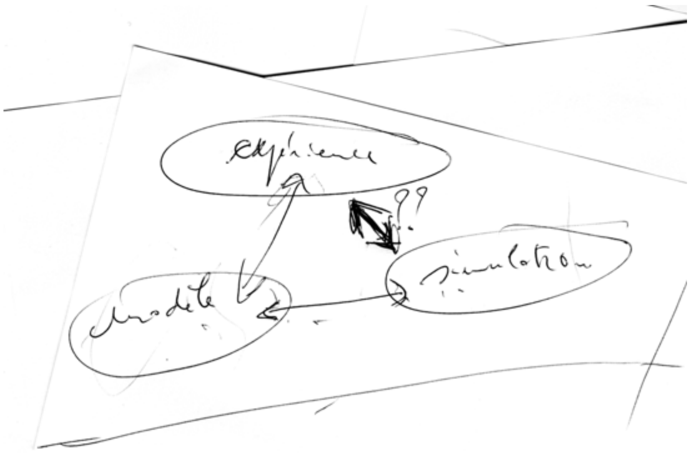
• Coutinho se limite à des expériences de Bernoulli, et a pour objectif principal d'introduire une étape intermédiaire avant la notion de probabilité d'un événement, qu'elle nomme « pré-probabilité ». Cette étape s'appuie sur des modélisations des expériences étudiées par des tirages d'un jeton d'une urne contenant des jetons de deux couleurs.

DEUXIEME PARTIE

Analyse de la tâche « simulation d'une expérience aléatoire »

Nous allons dans le chapitre 6 présenter deux simulations informatiques d'un jeu, assez connu dans l'enseignement secondaire sous le nom de « jeu du lièvre et de la tortue » : une simulation avec un *langage impératif*, et une simulation avec un tableur. Nous ferons apparaître que ces simulations ne sont pas des simulations de la loi de probabilité de l'expérience.

Nous proposerons alors, dans le chapitre 7, quelques définitions (modèle, expériences équivalentes, simulations), qui nous permettront de préciser notre problématique concernant les liens entre les trois sommets du triangle expérience / modèle / simulation ; nous pourrons alors faire apparaître que d'une part il y a des invariants dans les tâches de modélisation et de simulation d'une expérience aléatoire, et que d'autre part déterminer une simulation et une modélisation d'une expérience aléatoire sont deux tâches de natures différentes .



Dans le chapitre 8, nous préciserons les articulations entre simulation et modélisation d'une même expérience aléatoire, et réexaminerons en particulier le point de vue de Coutinho sur les simulations (2001).

Dans le chapitre 10, nous reviendrons sur l'aspect algorithmique d'une simulation informatique, en évoquant quelques problèmes didactiques posés pour l'enseignement de l'algorithmique-programmation.

Enfin dans le chapitre 11, nous commenterons la manière dont sont présentées les simulations informatiques d'expériences aléatoires dans des documents qui accompagnent les nouveaux « programmes 2010 ».

CHAPITRE 6

Etude d'un exemple

1. Le jeu du lièvre et de la tortue

Nous allons dans ce paragraphe présenter une simulation d'un jeu : « le jeu du lièvre et de la tortue ». L'étude que nous mènerons servira d'exemple ou d'illustration dans les chapitres suivants.

<u>1.1. Description du jeu et algorithmes de simulation</u>

> *Un lièvre et une tortue sont sur la même ligne de départ ; gagnera le premier qui franchira la ligne d'arrivée. Le lièvre atteint cette ligne en un bond, alors que la tortue ne l'atteint qu'après avoir fait six pas.*
> *On lance un dé. Si le six sort, le lièvre fait un bond et a gagné, sinon, la tortue avance d'un pas.*
> *<u>Question</u> : qui a le plus de chances de gagner ? Peut-on quantifier les chances de gagner du lièvre ?*

On veut simuler 100 parties du jeu du lièvre et de la tortue pour calculer la fréquence de l'événement « le lièvre gagne ».

• <u>Première étape</u> : faisons l'hypothèse que le dé est un dé cubique équilibré. Lancer le dé consiste donc à tirer un entier entre 1 et 6 selon une loi équirépartie, par exemple par une instruction telle que « Alea(1,6) » sur une calculatrice, si l'on programme la situation, ou en tirant *au hasard* une carte parmi l'as, le 2, le 3, le 4, le 5 et le 6 de pique d'un jeu de 52 cartes, les 6 cartes ayant été mélangées.

• <u>Deuxième étape</u> : si le 6 ne sort pas, la tortue avance d'un pas ; il faut imiter cela : par un compteur T du nombre de pas parcouru par la tortue. Au départ T = 0 puisque la tortue est encore sur la ligne de départ.
Si le 6 sort, le lièvre fait un bond, donc on peut compter le nombre L de bonds fait par le lièvre. Dans la mesure où dès qu'il fait un bond, le lièvre gagne, on pourrait se passer de cette

variable, mais nous préférons conserver le rôle symétrique des deux acteurs, ce qui permettra de définir facilement le test d'arrêt du jeu.

Au départ L = 0.
La partie est finie si T = 6 ou L = 1.
D'où :
• Algorithme de simulation pour une partie :

0 dans L et 0 dans T
répéter tant que L ≠ 1 et T ≠ 6
si Alea(1,6) = 6 alors 1 dans L sinon rajouter 1 à T
fin de répéter tant que

Remarquons que cet algorithme peut être considéré comme représentant une définition codée du jeu, c'est à dire la définition de l'expérience aléatoire qui nous intéresse ici :

L et T doivent faire des pas ; au début du jeu, ils n'ont pas encore fait de pas ;
tant que L n'a pas fait un pas et que T n'en a pas fait 6 :
on lance un dé cubique équilibré, et si le 6 sort, L fait un pas, sinon T fait un pas.

• Algorithme de simulation de 100 parties :
On veut calculer la fréquence de l'événement « le lièvre a gagné ». On crée un compteur X des parties gagnées par le lièvre. Au départ, X = 0

0 dans X
répéter 100 fois
algorithme pour une partie
si L=1 alors ajouter 1 à X
fin de répéter
calculer X/100

d'où l'algorithme complet :

0 dans X

répéter 100 fois

  0 dans L et 0 dans T

répéter tant que L ≠ 1 et T ≠ 6

    si Alea(1,6) = 6 alors 1 dans L sinon rajouter 1 à T

fin de répéter tant que

  si L=1 alors ajouter 1 à X

fin de répéter

afficher X/100

<u>1.2. mise en œuvre de l'algorithme « à la main »</u>

Cet algorithme peut être « exécuté à la main » en lançant un dé, on réalise alors 100 fois l'expérience aléatoire telle qu'elle est définie par la règle de jeu ; il peut être exécuté en remplaçant le lancer du dé par le tirage *au hasard* d'une carte comme décrit plus haut, l'exécutant remplissant au fur et à mesure un tableau de cent lignes du type. Réaliser ce tirage *au hasard* suppose la mise en place de conditions expérimentales : l'opérateur doit ne doit pas tricher, les cartes doivent avoir des dos identiques, les cartes doivent être « battues », etc.

<table>
<tr><td></td><td>déroulement de la partie</td><td>X = 0 puis :</td></tr>
<tr><td>partie 1</td><td><table><tr><td>lancer n° :</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>T = 0 puis :</td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>L = 0 puis :</td><td></td><td></td><td></td><td></td><td></td><td></td></tr></table></td><td></td></tr>
<tr><td>partie 2</td><td><table><tr><td>lancer n° :</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>T = 0 puis :</td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>L = 0 puis :</td><td></td><td></td><td></td><td></td><td></td><td></td></tr></table></td><td></td></tr>
<tr><td>…..</td><td></td><td></td></tr>
<tr><td>…..</td><td></td><td></td></tr>
<tr><td>partie 100</td><td><table><tr><td>lancer n° :</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>T = 0 puis :</td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>L = 0 puis :</td><td></td><td></td><td></td><td></td><td></td><td></td></tr></table></td><td></td></tr>
<tr><td>calcul de la fréquence de l'événement « le lièvre a gagné » :</td><td></td><td>diviser la valeur de la case qui est au dessus par 100</td></tr>
</table>

Ce tableau, d'après la remarque ci-dessus, peut être utilisé si l'on joue « pour de vrai » avec un dé.
Cependant l'organisation des résultats lorsqu'on joue « pour de vrai » peut être différente : on fait d'abord les 100 parties, en « cochant », au fil des parties, que le lièvre a gagné une nouvelle partie (par exemple en faisant un bâton par partie gagnée) et ce n'est qu'à l'issue des 100 parties qu'on fait le total du nombre de parties gagnées par le lièvre.
Ici, le nombre de parties « déjà gagnées par le lièvre » se fait « en temps réel ». Il est clair que ce choix n'est pas par hasard : il est nécessaire si l'on veut programmer la simulation dans un langage de programmation impératif, tel que le langage dont dispose chaque lycéen avec la calculatrice qu'il a dans son cartable.

L'algorithme mis en œuvre à la main **en lançant le dé supposé équilibré** est la répétition 100 fois de l'expérience « une partie du jeu du lièvre et de la tortue ».
L'algorithme mis en œuvre **en tirant au hasard un carton parmi 6 cartons numérotés de 1 à 6**, est une simulation de cette répétition 100 fois de l'expérience « une partie du jeu du lièvre et de la tortue », ou ce qui revient au même, 100 répétitions de la simulation d'une partie.
Cela revient au même, **si l'on accepte l'hypothèse que les tirages des nombres pseudo-aléatoires sont des expériences aléatoires indépendants, hypothèse que nous ferons dans toute la suite.**[48]

1.3. mise en œuvre de l'algorithme sur une calculatrice

Cet algorithme peut être aisément implémenté sur une calculatrice. Par exemple sur TI 82 :

```
0 STO X
For i, 1, 100
0 STO L : 0 STO T
While L ≠ 1 and T ≠ 6
If Alea(1, 6) = 6 then L+1 STO L else T+1 STO S End
End
If L = 1 then X+1 STO X End
Disp X/100
```

[48] Ces nombres pseudo-aléatoires sont les termes d'une suite définie par récurrence. Nous admettons que la définition de cette suite permet de faire cette hypothèse d'indépendance, et l'hypothèse que les nombres calculés suivent une loi uniforme.

C'est une traduction mot à mot de l'algorithme décrit en 1.1., à la seule différence de l'introduction du *compteur* i, qui est un choix fait par les concepteurs du langage de programmation de la calculatrice (et d'autres langages impératifs) pour implémenter les boucles où le nombre de répétition est connu.
L'exécution de ce programme produit la simulation de 100 parties du « jeu du lièvre et de la tortue », au même titre que l'exécution à la main de l'algorithme décrit en 1.1., puisque c'est la mise en œuvre de ce même algorithme.

Remarquons que la difficulté de cette simulation <u>à partir du moment où les élèves adhèrent à l'idée que la calculatrice peut produire du hasard analogue à celui du dé</u> , ne réside pas dans la simulation du lancer du dé, mais dans le test d'arrêt d'une partie : il y a là une question de logique qui ne va pas de soi ; la partie s'arrête si L= 1 ou si T = 6 (le ou étant ici exclusif, mais là n'est pas le problème) ; donc la partie se poursuit si : non (L=1 ou si T = 6).
La négation des connecteurs logiques « ou » et « et » n'est plus, dans les programmes des années 1990 et 2000, un objet d'apprentissage en tant que tel [49]. L'acquisition des règles de négation doit se faire « quand on en a besoin » préconisent les programmes. Ce peut être un avantage inattendu pour cette acquisition que d'avoir à écrire des simulations.
Nous reviendrons dans le chapitre 11 sur la programmation de ce jeu.

<u>1.4. Cet algorithme peut aussi être implémenté sur tableur.</u>

Il est un fait que les tableurs n'ont pas été conçus pour programmer des algorithmes itératifs : ce sont à l'origine des outils de gestion (stocks, comptabilité etc.), auxquels on a rajouté beaucoup de fonctionnalités qui en font, dans une certaine mesure, un outil efficace pour faire du calcul statistique.
Mais le tableur ne permet pas de répéter un calcul « tant que » une condition n'est pas satisfaite : ce qu'on peut faire, c'est répéter (on dit : recopier ou « tirer ») une formule <u>sur un nombre défini à l'avance de lignes</u> (ou de colonnes). L'avantage du tableur est qu'il ne « plante » pas si on lui demande une opération impossible, comme par exemple une division par zéro lorsqu'on fait une feuille de calcul pour l'algorithme d'Euclide par exemple : il signale le problème par un code dans la ou les cellules dont il n'a pas pu calculer la valeur.

Comment dans le jeu du lièvre et de la tortue traduire l'algorithme avec un tableur ?

---

[49] mais semble revenir sans les programmes des années 2010.

La simulation du lancer de dé ne pose aucun problème :

• Lorsque le tableur ne possède qu'une fonction aléatoire uniforme « Alea() » sur [0 ; 1], la formule « Ent(Alea()*6+1) » [50] est une formule que les élèves semblent comprendre, et retiennent assez bien (elle est introduite en classe de seconde) ; il y a pourtant une variable aléatoire non discrète en jeu.

• Les dernières versions d'*Excel* et d'*Open Office* possèdent la commande directe « Alea.entre.bornes(a,b) » qui donne un entier aléatoire entre les entiers a et b, selon une loi uniforme.

Ce qui pose problème ici, c'est le nombre de fois où l'on doit simuler le lancer de dé.

1.4.1. Une première solution, non satisfaisante : on lance six fois le dé

Cette solution est « qui peut le plus peut le moins », donc : on simule six lancers (au bout desquels on est assuré, qu'il y aura un gagnant), quitte à « lancer pour rien ». C'est la solution proposée dans les documents d'accompagnement de la classe de seconde (cf. [1] de la bibliographie)

Il y a là un problème qu'on ne saurait évacuer trop vite : il y a des simulations inutiles du lancer de dé, dont on ne tient pas compte. Est-ce que cela change « les chances » ? C'est une question qui dérangeait aux XVIIe et XVIIIe siècle dans la solution du problème des parties (Pascal, Fermat, puis d'Alembert), qui consistait à considérer qu'on poursuit la partie, même lorsque l'un des joueurs a gagné. Elle dérange toujours.

Tentons de reprendre l'algorithme décrit en 1.1.

Si l'on conçoit de traiter la simulation de la façon suivante : une partie dans un paquet de cellules situées sur une même ligne, la traduction de la répétition de 100 parties se fait tout à fait classiquement en tirant ce paquet de cellules vers le bas.

Il faut après chaque partie, actualiser le score du lièvre ; cela peut se faire en colonne B, sur la ligne k (i.e. dans la cellule Bk) si la ligne k correspond au numéro de la partie qui vient d'être jouée : on ajoute 1 au score antérieur, qui se trouve dans la cellule B(k-1) immédiatement au dessus si le lièvre a gagné la partie, on garde ce score sinon. Cela se traite classiquement avec la commande : = SI (*condition* ; B(k-1) + 1 ; B(k-1)).

[50] « Ent » note ici la fonction partie entière.

cellule B(k-1) cellule Bk

	A	B	C	D	E
1					
…					
k-1					
k		= SI (*condition* ; B(k-1) + 1 ; B(k-1))			
…					

La difficulté réside dans la construction d'une « bonne » condition : il s'agit de déterminer qui a gagné la partie de la ligne k.

Les solutions relèvent d'astuces.

Par exemple : on lance 6 fois le dé ; si le 6 ne sort pas, la tortue avance d'un pas. Si l'on considère qu'un bond du lièvre vaut 6 pas de tortue, on peut décider d'un indicateur calculé à partir de chaque lancer de dé : cet indicateur vaut 1 si c'est la tortue qui avance et 6 si c'est le lièvre, ce qui donne à chaque lancer (simulé) :

« SI (Alea.entre.bornes(1 ;6) = 6 ; 6 ; 1) ».

On calcule ainsi 6 indicateurs dans les cellules Ck, Dk, Ek, Fk, Gk et Hk, chacuns égaux à 1 ou à 6.

Une partie est gagnée par la tortue si ces 6 indicateurs sont à 1. Leur produit vaudra alors 1 et ne vaudra 1 que dans ce cas.

D'où la *condition* cherchée : NON(Ck * Dk*Ek * Fk * Gk * Hk = 1).

Finalement la feuille de calcul est construite par :

• 0 en B2 (le score de départ du lièvre est nul. Les parties commencent à partir de la ligne 3.

• en C3, D3, E3, F3, G3 et H3 : la formule

« =SI(ENT(ALEA() * 6 + 1)=6;6;1) »

• en B3 : la formule

« =SI(NON(C3*D3*E3*F3*G3*H3=1);B2+1;B2) »

• une recopie sur les 99 lignes suivantes des cellules B3 à H3.

• enfin en B104 :

« =B102/100 »

On obtient par re-calcul de la feuille (touche F9 sous Window) :

0,67 ; 0, 66 ; 0,68 ; 0,65 etc

La répétition d'un grand nombre de ces simulations par appui sur F9 ne demande que l'effort de noter le résultat, les re-calculs étant quasi instantanés.

SimulationLièvre.xls

	A	B	C	D	E	F	G	H	I
1	partie n°	Lièvre	lancer du dé n°1	n°2	n°3	n°4	n°5	n°6	
2		0							
3	1	1	1	6	1	1	1	1	
4	2	2	6	1	1	1	1	1	
5	3	2	1	1	1	1	1	1	
6	4	3	1	6	1	1	1	1	
7	5	3	1	1	1	1	1	1	
8	6	4	6	1	1	1	1	1	
9	7	4	1	1	1	1	1	1	
10	8	5	1	6	6	1	1	1	
11	9	6	1	1	1	6	6	1	
12	10	6	1	1	1	1	1	1	
13	11	7	6	1	6	1	1	1	
14	12	8	6	1	1	1	1	1	
15	13	9	1	1	1	6	1	6	
16	14	9	1	1	1	1	1	1	
17	15	10	1	1	6	1	1	1	
18	16	11	1	6	1	1	1	1	
19	17	11	1	1	1	1	1	1	
20	18	12	1	1	1	1	1	6	
21	19	13	6	1	1	1	1	1	
22	20	14	1	1	6	1	6	1	
23	21	14	1	1	1	1	1	1	
24	22	15	6	1	1	1	1	6	
25	23	15	1	1	1	1	1	1	
26	24	16	6	1	1	1	1	1	
27	25	17	1	1	1	1	6	1	
28	26	18	6	1	1	1	1	1	
29	27	19	1	1	6	1	1	1	
30	28	20	1	1	1	6	1	1	
31	29	21	6	1	1	6	1	6	
32	30	21	1	1	1	1	1	1	
33	31	22	1	1	1	1	6	6	
34	32	23	1	6	1	1	1	1	
35	33	24	6	1	1	1	1	6	
36	34	25	1	1	1	1	6	1	
37	35	25	1	1	1	1	1	1	
38	36	25	1	1	1	1	1	1	
39	37	25	1	1	1	1	1	1	

99	97	65	1	1	1	1	1	1	
100	98	66	1	1	1	1	1	6	
101	99	67	1	6	1	6	1	1	
102	100	68	1	1	1	1	1	6	
103									
104	fréquence	0,68							
105									
106									
107									

Remarquons que la ligne Bk a été définie afin de traduire l'algorithme défini en 1.1., algorithme défini pour une programmation dans un langage impératif : si l'on veut dans un tel langage organiser efficacement un cumul, ce cumul se fait « en temps réel », afin d'éviter de consommer de l'espace mémoire en sauvegardant les données à cumuler.

En revanche, avec le tableur, on peut prévoir d'indiquer sur chaque ligne qui a gagné (1 pour le lièvre, 0 pour la tortue) ; la formule « = SI (*condition* ; B(k-1) + 1 ; B(k-1)) ». est alors remplacée par : « = SI (*condition* ; 1 ; 0) ».

Le score final du lièvre se fait alors en calculant la somme des cellule de la colonne B, et comme c'est la fréquence que nous voulons calculer, il convient de mettre en B104 la formule : « =MOYENNE(B3 : B102) ».

Cette feuille de calcul correspond alors non pas à l'algorithme décrit en 1.1 (présenté dans la colonne de gauche du tableau qui suit), mais à l'algorithme présenté dans la colonne de droite où « score » est un tableau de longueur 100.

0 dans X répéter 100 fois 0 dans L et 0 dans T répéter tant que L ≠ 1 et T ≠ 6 si Alea(1,6) = 6 alors 1 dans L sinon rajouter 1 à T fin de répéter tant que si L=1 alors ajouter 1 à X fin de répéter calculer X/100	mettre score à 0 ; mettre 0 dans i répéter 100 fois 0 dans L et 0 dans T répéter tant que L ≠ 1 et T ≠ 6 si Alea(1,6) = 6 alors 1 dans L sinon rajouter 1 à T fin de répéter tant que si L=1 alors mettre 1 dans score(i) augmenter i de 1 fin de répéter calculer la moyenne des valeurs du tableau score

On remarquera que lorsque la traduction de « répéter 100 fois » se fait avec une boucle du type « For », cette boucle comporte un compteur, le compteur dont on a justement besoin pour actualiser le tableau « score » ; par conséquent la différence entre le programme obtenu par l'un ou l'autre de ces algorithmes ne réside pas dans ce compteur « i » de l'algorithme de droite, mais dans <u>l'espace mémoire</u> que nécessitera l'une ou l'autre des exécutions de ces programmes, du fait du tableau « score ».

On peut ici remarquer que la solution du cumul « sur place » nécessite des connaissances en programmation, concernant :

• l'instruction d'*affectation*,

• la valeur d'une *variable informatique* n'est pas la même au fil de l'exécution du programme
• l'interprétation que l'on peut faire de cette valeur n'est pas la même en différents points du programme.

1.4.2. Une solution « sans coup pour rien »

Au lieu d'avoir une conception d'ensemble du jeu : « tant que … , lancer le dé », on peut en avoir une vision plus locale « lancer le dé : si sort le 6 … sinon lancer le dé », ce qui permet, en imbriquant six conditionnelles de simuler une partie « sans coup pour rien » en une formule, au demeurant assez indigeste :

= SI(ALEA.ENTRE.BORNES(1;6) = 6 ; 1 ; SI(ALEA.ENTRE.BORNES(1;6) = 6 ; 1 ; SI(ALEA.ENTRE.BORNES(1;6) = 6 ; 1 ; SI(ALEA.ENTRE.BORNES(1;6) = 6 ; 1 ; SI(ALEA.ENTRE.BORNES(1;6) = 6 ; 1 ; SI(ALEA.ENTRE.BORNES(1;6) = 6 ; 1 ; 0) ,

Cette formule rend 0 si lors de 6 tirages aléatoires d'un entier entre 1 et 6, le six n'est jamais sorti : la tortue a gagné ; cette formule rend 1 dès que le 6 sort lors d'un tirage aléatoire (six tirages au plus) : c'est le lièvre qui a gagné.
Le tirage « Alea.entre.bornes(1 ; 6) » n'est fait que si le 6 n'est pas déjà sorti.

La feuille de tableur ainsi constituée ne correspond à aucun des deux algorithmes présentés dans le tableau du paragraphe 1.4.1.
La différence ne se situe pas dans la gestion de la répétition 100 fois de la simulation d'une partie, mais dans la simulation d'une partie.
L'algorithme présenté en 1.1 était fondé sur une vision globale du jeu :

> *Au départ, personne n'a fait de pas.*
> *On joue tant que personne n'a gagné.*
> *Le lièvre gagne s'il a fait un pas, la tortue gagne si elle a fait 6 pas.*
> *On lance un dé cubique : si le 6 sort, le lièvre fait un pas, sinon la tortue fait un pas.*

L'algorithme qui préside à la feuille de calcul présentée dans ce paragraphe est fondé sur une présentation on pourrait dire « minute par minute » d'une partie :

> *On lance une première fois le dé : si le 6 sort, le lièvre a gagné et la partie est finie, sinon, on lance une deuxième fois le dé : si le 6 sort, le lièvre a gagné et la*

partie est fini, sinon ...

etc.

Si après avoir lancé le dé une sixième fois le 6 n'est toujours pas sorti, c'est la tortue qui a gagné.

Ces deux descriptions du jeu correspondent à des descriptions « après coup » : elles sont en réalité extraites des algorithmes.
Cependant, elles définissent le même jeu.

1.5. En conclusion de ces mises en œuvre

• Les exécutions des programmes (calculatrice et tableur) que nous venons de définir peuvent être considérées comme la répétition d'une même expérience aléatoire, à condition d'accepter le générateur de nombres pseudo-aléatoires : nous faisons l'hypothèse que les générateurs utilisés simulent le tirage de nombres selon une loi uniforme sur un ensemble discret ou sur un intervalle, les résultats étant deux à deux indépendants.

• Ces expériences ont été construites sans concevoir au préalable le modèle probabiliste du jeu du lièvre et de la tortue, modèle qui consiste à associer leur probabilité à chacune des deux issues « le lièvre a gagné » et « la tortue a gagné » .

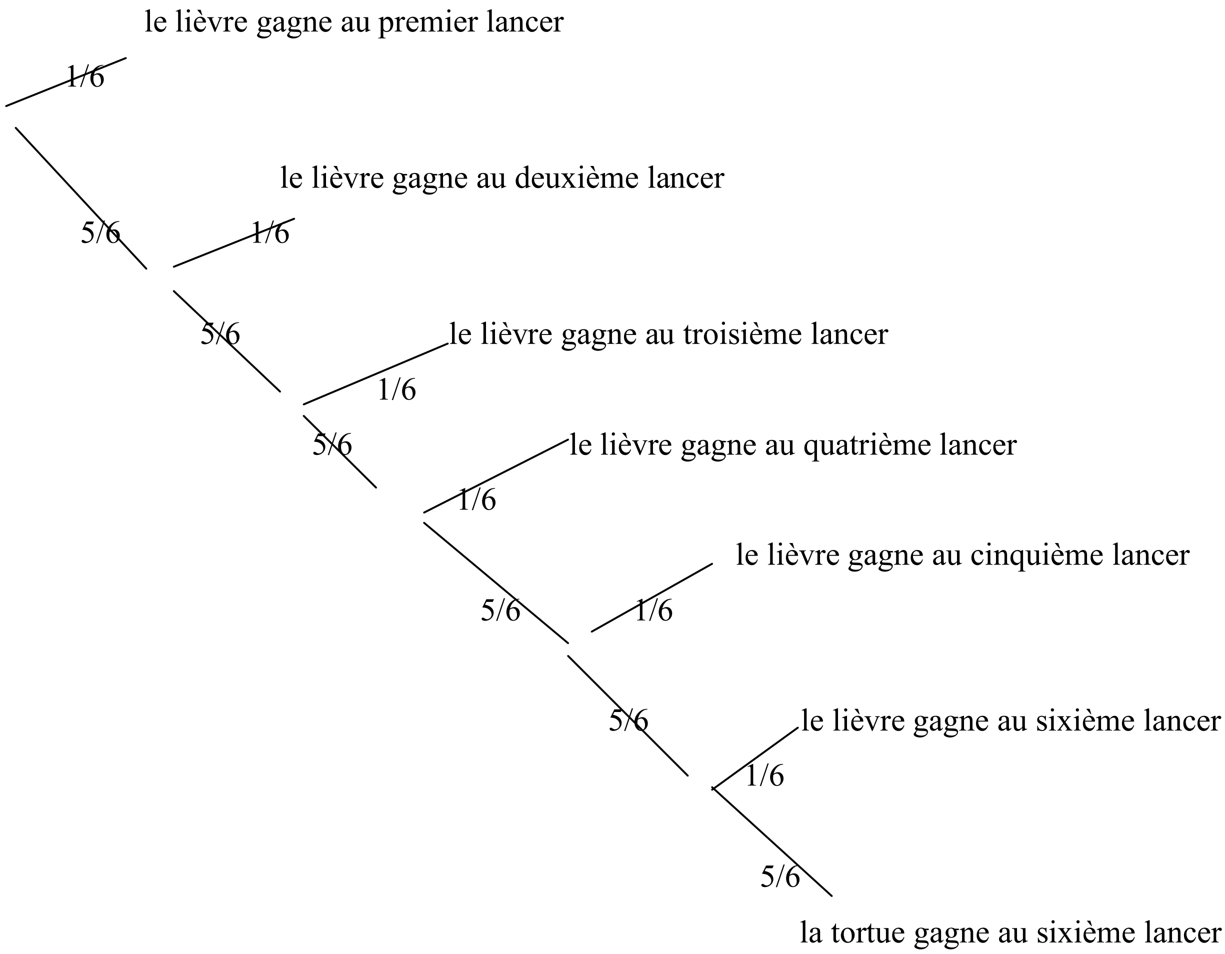
• La répétition 100 fois de ces expériences donne une distribution de fréquences significative : sous l'hypothèse d'un dé équilibré, le lièvre a environ deux chances sur trois de gagner (exactement $1 - (5/6)^6 \approx 0{,}6651$) ; nous présenterons ce calcul dans le paragraphe suivant.
Les simulations ne permettent pas de calculer de manière exacte cette valeur théorique ; elles permettent, en intégrant les fluctuations d'échantillonnage, d'en approcher.

• Les programmes présentés sont fondés sur des algorithmes qui décrivent le « jeu du lièvre et de la tortue », c'est à dire l'expérience aléatoire qui nous intéresse. Ces algorithmes, mis en œuvre « à la main » avec un dé *équilibré*, **sont en fait l'expérience aléatoire elle-même**.
Nous pouvons alors considérer que les autres mises en œuvre **sont des simulations de cette expérience**, informatiques (avec le programme pour la TI ou avec le tableur) ou non informatique (tirage au hasard d'un carton parmi 6).

• Enfin, nous avons remarqué qu'on peut présenter, pour une même expérience aléatoire, plusieurs « algorithmes », qui sont plus ou moins adaptés au langage de programmation dans lequel on veut les programmer. Nous mettons ici des guillemets pour marquer que le terme algorithme n'est peut-être pas ici tout à fait adapté : nous proposerons une autre expression au paragraphe 5 du chapitre 10.

1.6. Un modèle pour une partie au jeu du lièvre et de la tortue

Une partie est aisément représentée par un arbre de probabilités :



Les règles de calcul dans un arbre permettent de calculer la probabilité p qu'a la tortue de gagner : $p = (5/6)^6$ et donc la probabilité de gagner pour le lièvre est $1 - p = 1 - (5/6)^6$.
Evidemment, on peut également calculer la probabilité qu'a la tortue de gagner sans arbre, les événements ne pas sortir 6 au premier lancer, ne pas sortir six au deuxième lancer etc.. étant deux à deux indépendants, de probabilité 5/6.

Remarquons que cet arbre de probabilité est, si l'on ne tient pas compte des étiquettes, isomorphe à l'arbre que l'on aurait pu faire pour présenter le jeu par la description du paragraphe 1.4.2. :

On lance une première fois le dé : si le 6 sort, le lièvre a gagné et la partie est finie, sinon, on lance une deuxième fois le dé :

le 6 est sorti : le lièvre a gagné, la partie est finie

on lance le dé

le 6 n'est pas sorti, on relance le dé.

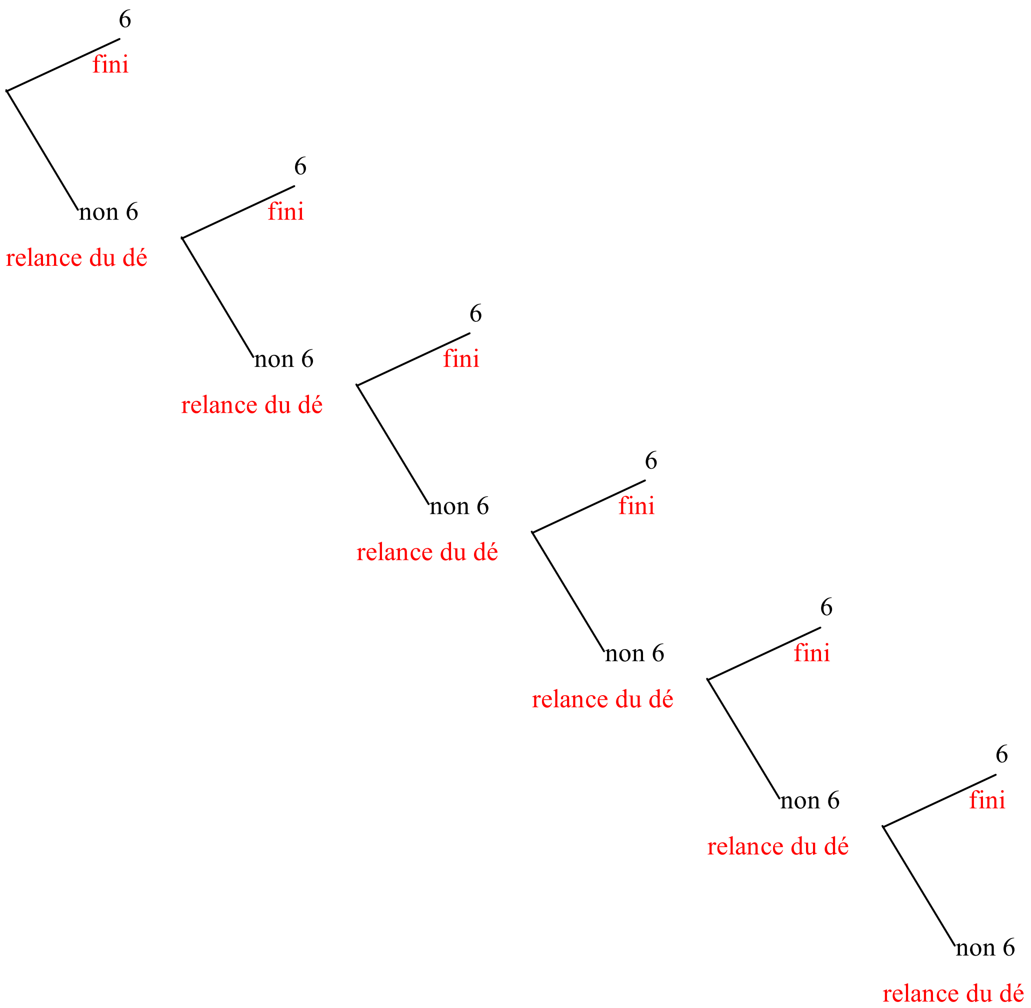
si le 6 sort, le lièvre a gagné et la partie est finie, …

etc.

On concatène plusieurs fois de suite l'arbre ci-dessus

Si après avoir lancé le dé une sixième fois le 6 n'est toujours pas sorti, c'est la tortue qui a gagné.

On obtient l'arbre :



2. Conclusions de cette étude

Cette étude nous a permis de faire ressortir un certain nombre de points que nous devrons approfondir :

• Nous avons présenté différents « algorithmes » décrivant la même expérience aléatoire.

On peut considérer que ces différents algorithmes peuvent rendre compte de diverses manières permettant de présenter cette expérience aléatoire.

Ils peuvent aussi être plus ou moins adaptés à un type de programmation (programmation dans un langage impératif ou programmation sur tableur).

• Nous avons présenté l'exécution de ces algorithmes (à la main ou par une machine) comme des *simulations* de l'expérience aléatoire d'origine. Cette affirmation était fondée sur le fait que ces algorithmes décrivent l'expérience elle-même, et que leur exécution « à la main » avec un dé n'est rien d'autre que la réalisation de l'expérience aléatoire d'origine.

• Nous avons vu que la répétition un grand nombre de fois de l'exécution des programmes informatiques traduisant ces algorithmes, est la répétition d'une expérience aléatoire qui produit des résultats qui paraissent compatibles, à la fluctuation d'échantillonnage près, avec les résultats qu'on peut observer en réalisant l'expérience « pour de vrai » [51].

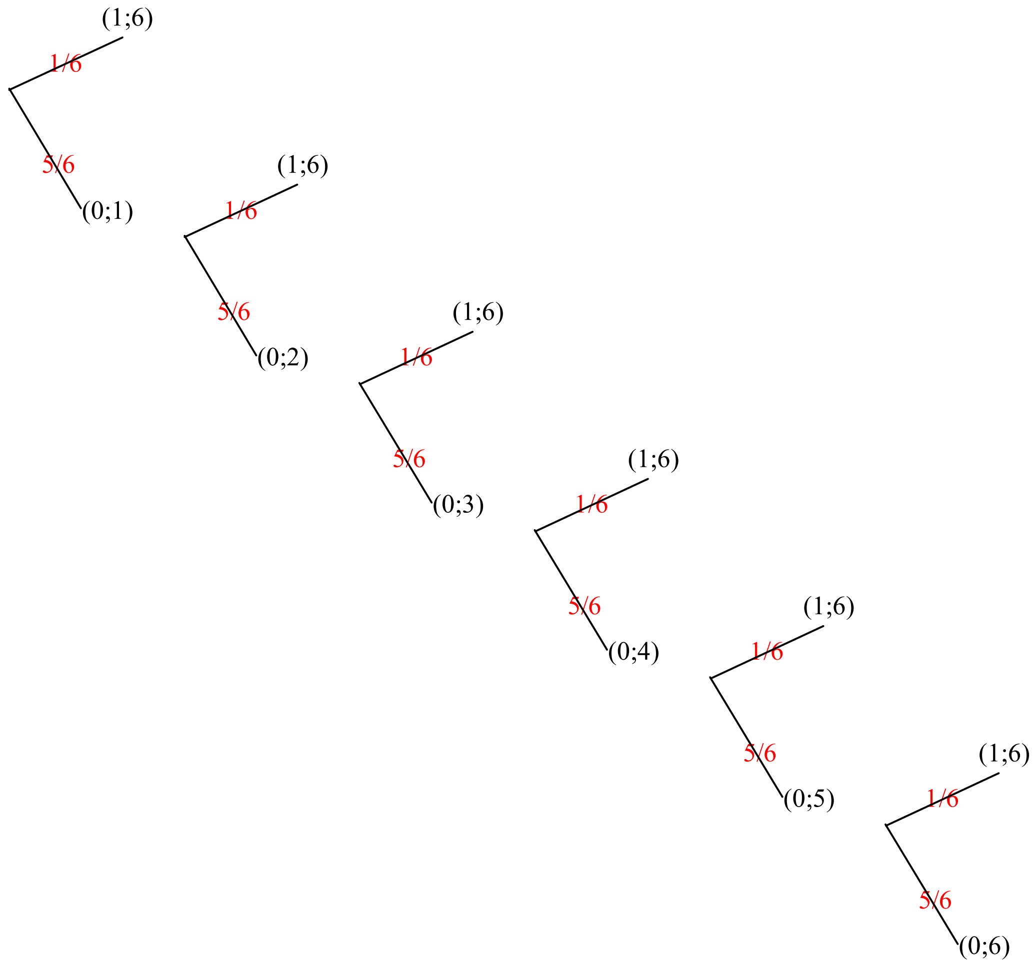
• Remarquons pour conclure ce chapitre que l'on peut déterminer la loi de probabilité de l'expérience aléatoire définie par exemple par le programme pour TI 82 indiqué au paragraphe 1.2. :

0 STO L : 0 STO T
While L ≠ 1 and T ≠ 6
If Alea(1, 6) = 6 then L+1 STO L else T+1 STO S End
End

Nous avons une expérience qui a deux issues possibles correspondant aux deux valeurs possibles du couple (L ; T) à l'issue de l'exécution du programme : (0 ; 6) ou (1 ; 0)

Si l'on construit l'arbre de probabilité associé à cette expérience, on obtient :

[51] Nous faisons réaliser 100 parties « pour de vrai » par nos élèves.



La probabilité de la feuille du bas étiquetée (0 ; 6) est de $(5/6)^6$; si l'on interprète cette feuille dans les termes « la tortue a gagné », la probabilité de l'événement contraire « le lièvre a gagné » correspondant à la réunion de toutes les feuilles étiquetées (1 ; 6) est de $1 - (5/6)^6$.

Les deux expériences aléatoires : le « jeu du lièvre et de la tortue » et l'exécution du programme ci-dessus ont des issues qui peuvent être mises en bijection, et qui ont même probabilité.

On constate que les deux expériences ont même loi de probabilité.

CHAPITRE 7

Des définitions

Il nous paraît nécessaire de préciser le sens que nous accorderons dorénavant à certains mots ou à certaines expressions.

1. Modèle probabiliste d'une expérience aléatoire

Dans le cadre de l'enseignement secondaire, un modèle probabiliste d'une expérience aléatoire est :

- soit un espace probabilisé (U, P(U), p) où U (ensemble fini) décrit les issues de l'expérience aléatoire, P(U) est l'ensemble des parties de U et modélise les événements liés à l'expérience aléatoire considérée, et p est une application probabilité définie sur P(U) ; cet espace probabilisé est complètement défini par la loi de probabilité de l'expérience : si l'univers contient k issues x_1, x_2, …, x_k, la loi de probabilité est le k-uplet (p_1, p_2, …, p_k) où p_i est la probabilité de l'issue x_i ; on attend en général des élèves qu'il présentent la loi de probabilité sous forme d'un tableau :

x_i	x_1	x_2	…	x_k
p_i	p_1	p_2	…	p_k

- soit un arbre de probabilité (les arbres de probabilité sont introduits dans les « programmes 2000 » en classes terminales) ; la composition d'un tel arbre et le calcul dans un arbre (qui met en œuvre, sans le dire, les probabilités conditionnelles et la formule des probabilités totales), sont maintenant bien codifiés dans l'enseignement secondaire,

- soit une loi de probabilité figurant au programme de terminale : loi équirépartie, loi binomiale, loi à densité continue (loi uniforme sur l'intervalle [a ; b] ou loi exponentielle sur IR^+). Pour une loi à densité, le choix d'une telle loi n'est pas à la charge de l'élève, il est en général déjà fait dans les énoncés des exercices ou des problèmes proposés.

- soit une combinaison simple de ce qui précède.

Les « programmes 2000 » de première (S, ES et L) mettent en place la notion d'espace probabilisé sous la forme d'un tableau donnant la loi de probabilité de l'expérience, comme nous venons de le décrire.
Les programmes de terminale, qui comportent les probabilités conditionnelles, permettent de rapporter presque tous les exercices proposés au baccalauréat à l'élaboration d'un arbre de probabilité décrivant la situation, suivi d'un calcul quasi mécanique s'appuyant sur cet arbre : la formule des probabilités totales (qui figure au programme et qui fonde ces calculs) n'est pas nécessairement évoquée dans les rédactions ; elle n'est pas attendue dans les exercices de baccalauréat.
Les probabilités à écrire sur les branches de l'arbre relèvent de l'une des lois énumérées ci-dessus, ou de données empiriques fournies avec le problème à résoudre.

Lorsqu'il y aura ambiguïté, nous le préciserons en parlant de « modèle probabiliste », sinon, nous dirons simplement « modèle » .

2. Expériences aléatoires élémentaires, expériences aléatoires composées

2.1. Expériences aléatoires élémentaires

Nous appellerons expérience aléatoire élémentaire les expériences telles que :
• le lancer d'un dé (équilibré ou non) à n faces,
• le lancer d'une pièce (équilibrée ou non),
• le tirage *au hasard* d'un jeton d'une urne dont on connaît la composition, ou toute autre expérience aléatoire consistant à tirer *au hasard* un objet dans un ensemble d'objets.
• une expérience de Bernouilli (expérience à deux issues, « succès » avec la probabilité p, et « échec » avec la probabilité $1- p$).

Lorsque la pièce ou le dé sont supposés *équilibrés*, ou lorsque le tirage est *au hasard* la loi de probabilité qu'on associe à l'expérience aléatoire est la loi équirépartie.
Dans le cas contraire, les énoncés donnent les indications « utiles » pour déterminer la loi de probabilité à associer à l'expérience aléatoire élémentaire en jeu.
Cependant rien dans les programmes n'interdit de travailler dans la classe avec un « vrai » dé, éventuellement pipé, pour essayer de déterminer la loi de probabilité à associer au lancer de

ce dé. Il n'existe pas, à notre connaissance, de matériel pédagogique pour mener un tel travail.

La simulation informatique d'une expérience aléatoire élémentaire peut se faire avec un appel d'une fonction aléatoire du langage de programmation ; si « Alea.entre.bornes(n ; p) » rend un nombre entier aléatoire entre les entiers n et p selon la loi équirépartie, et si « Alea() » rend un nombre aléatoire de l'intervalle [0 ; 1] selon la loi équirépartie, on simulera :

• le lancer d'un dé *équilibré* à n faces par « Alea.entre.bornes (1 ; n) », le lancer d'une pièce *équilibrée* par « Alea.entre.bornes (0 ; 1) ».

• une expérience de Bernouilli (expérience à deux issues, « succès » avec la probabilité p, et « échec » avec la probabilité 1- p) par « Alea() » ; un résultat inférieur à p sera associé à l'issue « succès », et un résultat supérieur à p sera associé à l'issue « échec ».

• Le lancer d'une pièce (équilibrée ou non) est une expérience de Bernouilli, et, dans le cas où la pièce n'est pas équilibrée, ce lancer peut se simuler comme ci-dessus.

• le tirage *au hasard* d'un jeton d'une urne contenant n jetons de k couleurs différentes, n_i jetons de la i – *ème* couleur ($\sum n_i = n$) par « Alea.entre.bornes (1 ; n) » ; un résultat inférieur à n_1 associé à l'issue « le jeton tiré est de la première couleur » ; entre $n_1 + 1$ et $n_1 + n_2$ le résultat sera associé à l'issue « le jeton tiré a la deuxième couleur », etc.

• le lancer d'un dé non équilibré peut se simuler comme le tirage *au hasard* d'un jeton d'une urne, dès que l'on connaît les *chances* de chacune des faces, c'est à dire la loi de probabilité que l'on peut associer au lancer de ce dé.

2.2. Expériences aléatoires composées

Une expérience aléatoire composée est une expérience aléatoire qui met en jeu la conjonction ou l'enchaînement d'au moins deux expériences aléatoires élémentaires. Ces expériences ne sont pas nécessairement indépendantes.

Le lancer de deux dés, le lancer de deux pièces sont des expériences composées ; le lancer peut être un lancer des deux objets simultanément, ou deux lancers successifs.

Le tirage (simultané) de deux jetons d'une urne est une expérience composée. Ce classement parmi les expériences aléatoires peut être discuté, il est justifié au paragraphe 8.2. de ce chapitre.

3. Modéliser une expérience aléatoire,

Modéliser une expérience aléatoire, c'est lui associer un modèle probabiliste (cf. paragraphe 1). Jusqu'à la quatrième partie de notre thèse, le verbe « modéliser » renverra toujours à la notion de « modèle probabiliste ».

Les « programmes 2000 » insistent particulièrement sur l'importance du choix du modèle : « l'énoncé vulgarisé » de la loi des grands nombres (chapitre 2, paragraphe 2) énonce que la « convergence » est une convergence dans le modèle ; le test d'adéquation de la loi équirépartie aux données expérimentales qui figure au programme des classes terminales et que nous détaillerons dans la quatrième partie, pose la question de la pertinence du choix de la loi de probabilité associée à l'expérience aléatoire étudiée.

Cependant, dans les exercices donnés au baccalauréat, ce qui reste à la charge de l'élève est en réalité peu important.

• La loi binomiale apparaît dans les exercices dans le cas d'une expérience qui peut se décrire en termes de « schéma de Bernouilli » : on répète n fois une expérience aléatoire qui n'a que deux issues, succès avec la probabilité p, et échec avec la probabilité $q = 1 - p$, les essais étant deux à deux indépendants. En général p est donné, ou bien p est la probabilité associée à une expérience aléatoire élémentaire, ou enfin p vient d'être calculé à l'issue des questions précédentes de l'exercice ou du problème.
Le plus souvent la variable aléatoire X égale au nombre de succès obtenus, dont un théorème du cours dit qu'elle suit la loi binomiale de paramètres n et p, est définie dans l'énoncé.

• Dans le cas d'un exercice mettant en jeu l'une des lois à densités continues du programme, le modèle de l'expérience aléatoire est donné ; l'élève doit déterminer, à partir de données fournies par l'énoncé, les valeurs des paramètres (la longueur de l'intervalle et / ou la constante égale à la densité pour une loi uniforme sur un intervalle, le paramètre « lambda » pour une loi exponentielle).

• Dans le cas où un arbre de probabilités est attendu, cette attente est en général inscrite dans l'énoncé (« on pourra faire un arbre de probabilité ») ; les événements utiles sont le plus souvent définis dans l'énoncé (« on notera P l'événement “ le test est positif “ et N

l'événement “ le test est négatif “).
Les probabilités des expériences aléatoires qui ne se décomposent pas en expériences élémentaires (voir le paragraphe précédent) sont également données dans l'énoncé : « une étude statistique indique que le test est positif chez 94 % des malades » ; il reste à la charge de l'élève l'interprétation en termes de probabilité de cette donnée mais l'impasse est faite sur tout ce qui se cache derrière l'étude statistique évoquée.

En définitive, la tâche de l'élève au baccalauréat consiste donc à interpréter les données de l'énoncé en termes de probabilités conditionnelles, puis à organiser son arbre de probabilités en conséquence (quels sont les événements étiquetés aux nœuds de profondeur 1 ?), et enfin à écrire les valeurs des probabilités sur les branches.
Pour calculer les probabilités demandées, l'élève utilise les règles de calcul dans un arbre de probabilité.

En classe de première, dans le seul cas des expériences aléatoires composées d'expériences aléatoires élémentaires, l'élève a à sa charge la définition de l'espace probabilisé : choix de l'univers et choix de la probabilité de chacune des issues composant cet univers [52].
Cela peut être par exemple le cas pour le lancer de deux dés, ou le tirage au hasard sans remise de deux jetons d'une urne dont on connaît la composition (ou le tirage de deux chaussettes d'un tiroir).
Dans ce dernier cas, on attend d'un élève de première une solution relevant d'une combinatoire élémentaire ; un élève de terminale pourra, lui, proposer une solution reposant sur un arbre de probabilités : les branches principales décrivent le premier tirage, et les branches secondaires décrivent le second tirage.

4. Expériences aléatoires équivalentes

Deux expériences aléatoires sont équivalentes si on peut leur associer un modèle probabiliste commun, à une bijection près sur l'ensemble des issues.
La définition de ces modèles passe éventuellement par des hypothèses, ou des réductions, sur ces expériences telles que : « le dé est équilibré », « la pièce ne peut retomber que sur l'une de ses deux faces ».

---

[52] Pour de telles situations, les élèves de terminale ont la faculté d'utiliser les arbres de probabilité, qui permettent d'esquiver la définition explicite de l'espace probabilisé associé à l'expérience.

Précisons la définition : pour une expérience aléatoire donnée E_1, un modèle probabiliste M_1 de cette expérience repose, explicitement ou non, sur un espace probabilisé (U_1, $\mathscr{B}_1$, p_1), qui est complètement déterminé, dans le cas où les issues sont dénombrables, par la donnée de l'univers U_1 décrivant ces issues, et par la donnée des images des événements élémentaires par l'application probabilité p_1 , la tribu des événements étant l'ensemble des parties de U_1 dans l'enseignement secondaire.
Considérons donc deux expériences aléatoires E_1 et E_2.
E_1 et E_2 sont des expériences aléatoires équivalentes si on peut construire deux modèles
$M_1= (U_1, \mathscr{B}_1, p_1)$ de E_1, et $M_2= (U_2, \mathscr{B}_2, p_2)$ de E_2
tels qu'il existe une bijection f de U_1 vers U_2 vérifiant : $p_1 = p_2 \circ f$. [53]
Nous définissons ainsi une relation d'équivalence sur les expériences aléatoires, ce qui justifie de parler d'expériences aléatoires équivalentes.

Dans la pratique, dans l'enseignement secondaire, la question de l'équivalence de deux expériences aléatoires se pose lorsque les univers sont finis. Dans ce cas, deux expériences aléatoires sont équivalentes si elles ont le même nombre d'issues et la même loi de probabilité associée à ces issues (à l'ordre des probabilités des issues près).
Et deux expériences aléatoires relevant de la loi équirépartie sont équivalentes si elles ont le même nombre d'issues.

Donnons quelques exemples :
• tirer au hasard un papier d'un sac contenant six papiers numérotés, lancer un dé équilibré, activer la fonction « EntAlea(1 ; 6) » d'une calculatrice ou la fonction « Alea.entre.bornes(1 ; 6) » d'un tableur sont des expériences équivalentes qui ont pour modèle probabiliste la loi équirépartie sur un ensemble à six issues.
• Lancer deux dés équilibrés simultanément ou lancer deux fois de suite un dé équilibré sont deux expériences équivalentes.
• tirer au hasard simultanément deux jetons d'une urne, ou tirer successivement et sans remise deux fois un jeton de cette même urne sont deux expériences équivalentes.
Remarquons que c'est, entre autres raisons, l'équivalence entre ces deux expériences qui nous

[53] Nous commettons pour simplifier un abus de notation : l'application probabilité p agit, non sur les éléments u de l'univers U, mais sur les événements de la tribu définie sur U. Au lieu d'écrire p({u}), probabilité de l'événement élémentaire {u}, nous écrivons p(u), comme si p était définie sur U.

a amenée à classer le tirage de deux jetons (sans remise) parmi les expériences composées .

5. Simulation

Une simulation d'une expérience aléatoire est :

• une expérience aléatoire équivalente à l'expérience aléatoire initiale,

• et une expérience effective, qui doit pouvoir être répétée et qui doit pouvoir être raisonnablement mise en œuvre par des moyens simples (tirage de jetons d'une urne par exemple), ou sur ordinateur (élaboration d'une feuille de calcul pour un tableur ou programmation dans un langage de programmation impératif par exemple).

Une simulation est conçue pour fournir des résultats expérimentaux (on parlera de données empiriques de la simulation).

Sous cette définition, on ne peut considérer comme le fait Schwartz (2011) que le lancer d'un dé équilibré est une simulation de la loi équirépartie sur {1 ; 2 ; … ; 6} puisqu'il ne s'agit pas, telle qu'elle l'entend, d'une expérience aléatoire effective, mais d'une expérience « virtuelle ».

6. Simuler une expérience aléatoire

Simuler une expérience aléatoire, c'est construire une simulation de cette expérience.

Nous avons présenté dans le chapitre précédent un exemple de construction de simulations du « jeu du lièvre et de la tortue ».

Nous avons dans un premier temps présenté l'exécution de l'un des programmes proposés, comme étant une simulation de l'expérience aléatoire initiale, sans justifier cette affirmation.

Nous avons conclu le chapitre sur la remarque que les expériences aléatoires que nous avions construites, ont même modèle probabiliste que l'expérience initiale.

Par conséquent, nous avons démontré que ces expériences obtenues par l'exécution des programmes, sont bien des simulations de l'expérience initiale, selon la définition que nous venons de proposer d'une *simulation* d'une expérience aléatoire.

Nous garderons dans la suite de notre travail ces définitions.

Remarquons que la définition d'une simulation « selon les programmes » est restrictive par rapport à la définition que nous en donnons : une simulation « selon les programmes » a, par construction de la simulation, même modèle probabiliste que l'expérience aléatoire d'origine, et lui est donc bien équivalente ; c'est donc une simulation au sens du paragraphe 5 de ce chapitre.

Mais une expérience aléatoire, pour mériter la qualité de simulation d'une expérience aléatoire « selon les programmes », doit avoir été construite à partir de la loi de probabilité de cette expérience. Selon ce point de vue, les programmes informatiques que nous avons présentés au chapitre précédent ne généreraient pas des simulations du jeu du lièvre et de la tortue.

7. Le problème de la preuve qu'une simulation est bien ce qu'elle prétend être

L'exemple du « jeu du lièvre et de la tortue » fait donc clairement apparaître un problème : nous avons construit une expérience aléatoire sans avoir associé un modèle probabiliste à cette expérience. Nous avons pu démontrer ensuite que cette expérience est bien une simulation du jeu parce qu'elle a même modèle probabiliste que lui.

Cependant, si nous effaçons cette preuve, il demeure que cette expérience est une simulation de l'expérience aléatoire, et que nous l'avons construite sans déterminer son modèle probabiliste.

Cet exemple n'est pas isolé : nous pouvons, par une démarche analogue, construire des programmes associés à des expériences aléatoires, pour lesquels nous postulons que leurs exécutions sont des simulations des expériences de départ.

Nous pouvons le faire, semble-t-il pour toutes les expériences aléatoires composées (en nombre déterminé ou non à l'avance) d'expériences aléatoires élémentaires qu'il est raisonnable d'envisager au lycée, y compris pour résoudre le « problème des tablettes de chocolat » que nous analyserons au paragraphe 3 du chapitre 9, problème pour lequel l'expérience aléatoire est composée d'un nombre indéterminé et non limité d'expériences aléatoires élémentaires : l'univers qui décrit les issues est donc infini.

Nous postulons, avons-nous dit, que les expériences ainsi construites sont bien des simulations, sans pourtant démontrer que l'expérience créée par l'exécution du programme informatique et l'expérience aléatoire initiale ont même modèle probabiliste.

Nous étions assez isolée au début de notre thèse à soutenir ce « postulat », nous appuyant en cela sur notre pratique : notre démarche pouvait être alors qualifiée d'introspective.
Nous avons ensuite constaté que dans les sujets de l'épreuve pratique expérimentale (cf. les annexes du chapitre 3), la simulation demandée précède la détermination de la loi de probabilité de cette expérience.
Nous constatons en 2010 et en 2011 que les exemples de simulations d'expériences aléatoires proposés dans divers documents accompagnant officiellement les programmes (cf. [1] et [2]) sont des algorithmes [54] qui n'ont pas été construits à partir de la loi de probabilité de ces expériences.
Ces documents sont aussi bien des documents accompagnant les programmes de probabilité, que des documents accompagnant la partie « algorithmique-programmation » des « programmes 2010 » de mathématiques.
Ces documents ne posent jamais la question de la preuve qu'il s'agit bien de simulations : ils déclarent que ce sont des simulations.
Ces documents ne présentent pas d'analyse ou de commentaires expliquant comment les programmes ou algorithmes ont été conçus.

Cette question est donc l'un des problèmes que nous devons tenter de résoudre : pourquoi un algorithme engendre-t-il une simulation d'une expérience aléatoire quant il n'a pas été construit à partir du modèle probabiliste de cet expérience ?

La question est d'importance, car de la réponse qu'on y apporte dépend l'utilisation que l'on peut faire des simulations informatiques d'expériences aléatoires : si une simulation « contient » le modèle probabiliste de l'expérience aléatoire, cette simulation apportera peu de choses à l'étude de l'expérience aléatoire d'origine ; elle pourra permettre une illustration de la loi des grands nombres ou de la notion d'intervalles de fluctuations (« programmes 2010 » de seconde) si on répète un grand nombre de fois la simulation, ce que permet facilement l'outil informatique.

[54] dans un souci d'allègement nous confondrons un programme informatique ou un algorithme, qui ne sont que du texte, de leurs exécutions (à la main ou par une machine), exécutions dont nous avons dit que c'est une expérience aléatoire, si au moins l'une des instructions fait appel à une fonction aléatoire.

8. Les hypothèses de modélisation

8.1. Un choix essentiel

On remarquera que les deux actions, modéliser et simuler, produisent un résultat : un modèle ou une simulation, en général parmi d'autres possibles.
Un modèle probabiliste et une simulations d'une expérience aléatoire sont en effet « adossés » à une ou des hypothèses, qui présideront à l'élaboration du modèle probabiliste ou de la simulation que l'on associera à cette expérience aléatoire, hypothèses que nous appellerons *hypothèses de modélisation*.

Ces hypothèses concernent les expériences élémentaires qui composent l'expérience globale.
Elles vont déterminer la loi de probabilité que l'on va pouvoir associer à chacune de ces expériences aléatoires élémentaires.
Ainsi pour le lancer d'un dé cubique :
• on peut faire l'hypothèse que le dé est équilibré, on associe alors au lancer de ce dé la loi équirépartie sur {1 ; 2 ; … ; 6}, et on pourra simuler le lancer de ce dé par un appel à « Alea.entre.bornes(1 ; 6) » comme indiqué au paragraphe 1.2. de ce chapitre.
• On peut faire une étude statistique sur la fréquence de sortie de chacune des faces lors d'un très grand nombre de lancers du dé, et être amené à faire l'hypothèse, en s'appuyant sur une étude de ces fréquences, que l'on peut associer au lancer de ce dé la loi (0,16 ; 0,20 ; 0,16 ; 0,16 ; 0,20 ; 0,12), et on pourra le simuler par l'appel de « Alea.entre.bornes (1 ; 100) » : un résultat de cet appel compris entre 1 et 16 sera associé à l'issue « le dé est tombé sur la face 1 » ; entre 17 et 36, le résultat sera associé à l'issue « le dé est tombé sur la face 2 », etc, comme indiqué au paragraphe 1.2. .

Ces *hypothèses de modélisation* ne sont pas tout à fait ce que Henry (1999) et 2001) et Parzsysz (2010), appellent les *hypothèses de modèle* : nous employons l'expression *hypothèses de modélisation* dans un sens plus restrictif ; il s'agit des hypothèses que l'on fait en tenant compte des conditions expérimentales (observées ou supposées), avant de déterminer un modèle probabiliste ou une simulation d'une expérience aléatoire, et ces hypothèses de modélisation vont porter sur les expériences élémentaires qui composent l'expérience : elles déterminent le modèle probabiliste de ces expériences élémentaires.

Les *hypothèses de modélisation* ne sont donc pas des hypothèses sur le modèle probabiliste associé à l'expérience globale (quand l'expérience aléatoire étudiée est composée), et ce modèle, pour l'expérience globale, reste à déterminer lorsque les *hypothèses de modélisation* ont été choisies.

Autrement dit, il ne s'agit pas, en choisissant les *hypothèses de modélisation*, de présenter un modèle probabiliste (une loi de probabilité) pour l'expérience globale à étudier, loi de probabilité pour laquelle on ferait l'hypothèse qu'elle modélise « correctement » cette expérience aléatoire (de telles hypothèses sont les *hypothèses de modèle* définies par Henry, cf. l'extrait ci-dessous), il s'agit de faire des hypothèses sur le modèle probabiliste que l'on associe <u>aux expériences aléatoires élémentaires qui composent l'expérience globale</u>.
Ces hypothèses permettront de construire une simulation, et, éventuellement, de construire un modèle probabiliste de l'expérience globale.

Pour Henry (2001), la loi de probabilité est l'une des *hypothèses de modèle* :

> « *(...) la probabilité d'un événement est définie comme la somme des probabilités des événements élémentaires (représentant les issues de l'expérience aléatoire) qui le constituent. Ces dernières font partie des* **hypothèses de modèle** *: estimées à partir de l'observation des fréquences stabilisées ou issues d'un calcul a priori à partir d'une hypothèse d'équiprobabilité* »

Cette hypothèse d'équiprobabilité évoquée par Henry pour mener un « calcul a priori » de la loi de probabilité, correspond précisément à l'*hypothèse de modélisation*, qui déterminera ce calcul des probabilités des événements élémentaires.

Pour une expérience aléatoire donnée, il y a, a priori, plusieurs choix d'*hypothèses de modélisation* possibles, comme nous l'avons illustré pour le lancer d'un dé.

Le choix des *hypothèses de modélisation* va déterminer une et une seule loi de probabilité pour l'expérience globale, loi de probabilité qui reste à calculer, au travers de calculs menés dans la théorie des probabilités prenant en charge le caractère composée de l'expérience : espace produit, arbres de probabilités ou image de la loi de probabilité par une variable aléatoire .

Ce choix des *hypothèses de modélisation* précède la détermination du modèle probabiliste de l'expérience aléatoire, mais est également nécessaire pour construire, selon la technique illustrée au chapitre 6 pour la simulation du « jeu du lièvre et de la tortue », une simulation de l'expérience aléatoire.

Ces *hypothèses de modélisation* peuvent être énoncées directement par une loi de probabilité : « on estime que la probabilité d'atteindre la cible est 0,8 », ou par une locution qui caractérise sans ambiguïté une loi de probabilité ; l'expression « le dé est équilibrée » est associée à la loi de probabilité équirépartie pour les faces du dé.

8.2. Les *hypothèses de modélisation* ne sont pas en général la loi de probabilité

Il est vrai que pour une expérience aléatoire élémentaire, on peut dire que les *hypothèses de modélisation* associées à cette expérience coïncident avec son modèle probabiliste, c'est à dire la loi de probabilité qu'on associe à l'expérience : dire qu'une pièce est **équilibrée** ou qu'un tirage d'un jeton d'une urne se fait **au hasard**, c'est faire l'hypothèse d'une loi équirépartie sur un ensemble d'issues.

Certaines expériences aléatoires ont un statut variable selon le niveau de la classe où on les considère : élémentaires ou composées.
De ce fait, déclarer que trois jetons sont tirés simultanément **au hasard** d'une urne qui contient n jetons, est ambigu quant aux hypothèses de modélisation induites par cette déclaration :

- pour un élève qui a quelque connaissance en analyse combinatoire, l'expérience peut être considérée comme élémentaire, et la déclaration d'un tirage *au hasard* revient à associer à l'expérience globale la loi équirépartie sur un univers à $n(n-1)(n-2)/6$ éléments.
- Pour un élève qui n'a pas de connaissance en analyse combinatoire, cela revient, s'il est convaincu de l'équivalence entre le tirage simultané de trois jetons et trois tirages successifs sans remise, à faire trois hypothèses de modélisation, une pour chaque tirage : cela amène alors à associer la loi équirépartie sur un univers à n éléments pour le premier tirage, à associer la loi équirépartie sur un univers à n-1 éléments pour le second tirage, etc.

La loi de probabilité à associer à cette expérience demandera pour un élève de première (« programmes 2000 ») quelques calculs.
La simulation de l'expérience « tirer trois jetons au hasard » passera alors par la simulation de chacun des trois tirages, grâce à un appel à « Alea.entre.bornes(1 ; n) », puis

« Alea.entre.bornes(1 ; n-1) », etc.

Mais la simulation de l'expérience aléatoire globale demande de prendre en compte non seulement le nombre de jetons dans l'urne avant chaque tirage, mais également sa composition précise à l'issue du tirage précédent (quels sont nommément les jetons qui restent). La simulation est délicate dans la mesure où les langages de programmation utilisés actuellement dans l'enseignement secondaire n'offrent pas, à notre connaissance, le tirage au hasard de k objets parmi n objets numérotés en « fonction primitive » du langage ; c'est pourquoi pour simuler un tel tirage, nous sommes amenés à considérer ce tirage comme une expérience composée d'expériences élémentaires : tirer un jeton d'une urne dont la composition est connue.

C'est en réalité pour cette raison que nous avons classé les tirages au hasard simultanés de jetons d'une urne parmi les expériences aléatoires composées.

Nous n'avons pas observé dans les exercices et problèmes de simulations proposés à l'étude en lycée de situations aléatoires comportant de tels tirages simultanés.

On peut cependant peut-être envisager avec les nouveaux programmes de mathématiques de telles situations.

On voit sur cet exemple du tirage de trois jetons d'une urne, que les hypothèses de modélisation ne donnent pas immédiatement en classe de première selon les « programmes 2000 » le modèle probabiliste à associer à l'expérience aléatoire.

Cela vaut de manière générale pour les expériences composées : comme nous l'avons déjà dit, même si le choix des hypothèses de modélisation déterminent le modèle probabiliste de l'expérience globale, ce choix étant fait, la loi de probabilité de l'expérience globale reste à calculer, ce qui n'est pas a priori nécessaire si la tâche visée est la simulation de l'expérience globale.

L'exemple du « problème des tablettes de chocolat » que nous allons présenter au paragraphe 3 du chapitre 9 est particulièrement probant : nous proposerons une simulation informatique de l'expérience aléatoire en jeu (acheter des tablettes de chocolat tant que la collection d'images n'est pas complète), sans avoir jamais déterminé le modèle probabiliste de cette expérience dont l'univers est infini dénombrable.

8.3. Un quiproquo dans la définition d'une *simulation* ?

La multiplicité des choix des hypothèses de modélisation qui président à la détermination du modèle probabiliste et de la simulation que l'on peut associer à une expérience aléatoire, amène bien entendu une multiplicité des modèles et des simulations que l'on peut associer à cette expérience aléatoire.
Cela justifie l'insistance des « programmes 2000 » à souligner la relativité en quelque sorte du modèle (l'énoncé proposé de la loi des grands nombres fait ressortir que la « convergence » de la fréquence vers la probabilité est relative au modèle choisi). Cela justifie également que soit posée la problématique de l'adéquation du modèle choisi au travers de la problématique « adéquation du modèle équiréparti aux données expérimentales » qui figure dans les « programmes 2000 » de terminale.

Cependant, le modèle probabiliste n'est « relatif » que parce qu'il est lié à un choix des *hypothèses de modélisation* parmi d'autres choix possibles, mais une fois choisies ces *hypothèses de modélisation*, le modèle probabiliste est complètement déterminé.

Lorsque la définition de « *simuler une expérience aléatoire* » est donnée dans les programmes dans les termes : « simuler une expérience aléatoire, c'est choisir un modèle de cette expérience, puis simuler ce modèle », on peut comprendre la locution « choisir un modèle de cette expérience » comme signifiant : « choisir *les hypothèses de modélisation* de cette expérience », puisque dans le cas d'une expérience aléatoire élémentaire, cela revient à choisir le modèle de cette expérience aléatoire élémentaire.
La simulation d'une expérience aléatoire élémentaire peut alors se faire par exemple par le tirage d'un jeton d'une urne de composition connue, ou par l'appel d'une fonction aléatoire comme indiqué au paragraphe 2.1. ; la simulation de l'expérience globale se fait « en composant » les simulations des expériences aléatoires élémentaires. Il resterait à préciser comment se fait cette « composition ».

Cependant on trouve dans les « programmes 2000 » quelques occurrences d'une définition de la simulation voisine : « simuler une expérience, c'est simuler sa loi de probabilité » ce qui paraît sans équivoque sur la nécessité de déterminer d'abord un modèle probabiliste avant d'envisager de déterminer une simulation.

8.4. Le choix des hypothèses de modélisation

Le choix des hypothèses de modélisation est un préalable à l'élaboration d'un modèle probabiliste ou d'une simulation d'une expérience aléatoire.
A qui incombe ce choix ?

En classes de première et au baccalauréat, (pour les « programmes 2000 »), les choix des hypothèses de modélisation pour une expérience composée peuvent en général se ramener aux hypothèses de modélisation qu'il convient d'associer au lancer d'un dé ou d'une pièce **équilibrés**, ou au tirage d'un jeton dans une urne contenant un nombre fini de jetons présentant diverses caractéristiques (numéro, couleur), tirage **au hasard**. Presque toujours donc la loi induite par les hypothèses de modélisation est une loi équirépartie sur un nombre fini d'issues.

Les expériences aléatoires sur lesquelles sont faites ces hypothèsesde modélisation sont simulées sans difficulté sur ordinateur ou sur calculatrice (dans une feuille de calcul de tableur ou par un programme dans un langage impératif) par un appel à une fonction aléatoire du langage de programmation utilisé, comme nous l'avons montré au paragraphe 2.1..

Par contre, une difficulté peut résider, lors de la détermination du modèle probabiliste, dans la traduction correcte des hypothèses de modélisation en une loi de probabilité sur les issues retenues pour décrire l'expérience aléatoire globale. Nous allons présenter cette difficulté dans le paragraphe suivant.

9. imitation d'une expérience aléatoire

9.1. Issues physiques, issues intéressantes

Nous allons être amenée à définir, pour une expérience aléatoire, la notion d'*issues physiques*, et celle d'*événements intéressants* ou d'*issues intéressantes*.

Lorsque l'on veut définir les issues d'une expérience aléatoire (par exemple pour déterminer la loi de probabilité que l'on peut associer à cette expérience), on peut être amené à distinguer plusieurs points de vue. Ainsi lorsqu'on lance deux dés :
• si l'on observe les numéros sortis, on décrira a priori les issues par : « double 1 », « 1 et 2 », …, « 1 et 6 », « 2 et 3 », …….., « double 6 », soit 21 issues ;

• si l'on joue au « Monopoly », on décrira plutôt les issues par : « 2 », « 3 », …., « 12 », soit 11 issues correspondant aux valeurs possibles de la somme des points amenés ;
• si l'on joue à un jeu où l'on avance de l'écart entre le numéro sorti sur le dé bleu et 2 fois le numéro sorti sur le dé rouge, il faudra distinguer les deux dés, et les issues considérées seront peut-être « double 1 », « 1 sur le rouge et 2 sur le bleu », « 2 sur le rouge et 1 sur le bleu », etc., soit 36 issues.
Les issues considérées peuvent aussi être les issues « utiles » pour ce jeu hypothétique, c'est à dire ici : 0, 1, …. et 11, soit 12 issues.

Il s'agit bien toujours de la même expérience aléatoire qui consiste à lancer deux dés, mais les points de vue diffèrent selon les « finalités » de cette expérience.
Il y a cependant un ensemble d'issue qui est sous-jacent à chacun de ces points de vue : c'est l'ensemble dont nous pourrions dire qu'il est « au plus près » de l'expérience aléatoire, c'est à dire l'ensemble à 36 éléments que nous avons évoqué, qui peut se représenter par l'ensemble Ω des couples (1 ; 1), (1 ; 2), …, (1 ; 6), (2 ; 1), …et (6 ; 6).
A partir de ce choix de représentation des issues, on peut décrire toutes les situations définies ci-dessus. Avec ce choix d'univers Ω, les « issues de l'expérience » décrivant en première intention le point de vue lié à chacune de ces situations deviennent des *événements*.

Remarquons que chacune des issues de cet univers Ω est reliée de manière bi-univoque au résultat du lancer : une issue décrit complètement le résultat *physique* de l'expérience ; le couple (2 ; 3) indique non seulement que les numéros sortis sont un 2 et un 3, mais également sur quel dé est sorti le 3. Aucune des autres descriptions des issues présentées ci-dessus ne permet cette « reconstruction ». Nous dirons de Ω que c'est l'univers des *issues physiques* de l'expérience globale.

Les autres univers décrivent en fait les événements, liés à l'expérience, qui sont intéressants pour diverses raisons (peut-être pour calculer la probabilité de l'un de ces événements en particulier). Pour de tels événements, nous parlerons d'*événements intéressants*. Lorsqu'un univers est constitué pour rendre compte des *événements intéressants*, nous dirons que c'est l'univers des *issues intéressantes* de l'expérience globale.
Les *issues intéressantes* d'une expérience aléatoire sont en grande partie déterminées par la finalité assignée à cette expérience aléatoire : lance-t-on les deux dés parce qu'on joue au Monopoly ou parce que, jouant au « 421 », on a fait le choix de relancer deux des trois dés ?

9.2. La portée des *hypothèses de modélisation*

Si l'on reprend l'expérience du lancer de deux dés cubiques, en faisant maintenant comme *hypothèses de modélisation* que chacun des deux dés est équilibré, ces hypothèses vont avoir des répercussions sur la détermination de la loi de probabilité associée au lancer des deux dés, et sur la détermination de la simulation.

• Pour la simulation : c'est sur les *expériences élémentaires composant* l'expérience que l'on a fait des hypothèses, c'est à dire sur chacun des deux dés.
Les *hypothèses de modélisation* sur chacun des deux dés disent que, pour chaque dé, la loi de probabilité à associer au lancer du dé est la loi équirépartie sur un univers à 6 éléments. Le lancer de chaque dé pourra alors être simulé par l'appel « Alea.entre.bornes(1 ; 6) ».

• Pour la détermination de la loi de probabilité à associer au lancer de deux dés cubiques équilibrés : le choix de l'univers n'est évidemment pas indifférent pour mener le calcul de la loi de probabilité que l'on va associer à l'expérience aléatoire.
Les *hypothèses de modélisation* induisent des hypothèses sur les *issues physiques* de l'expérience globale, et non des hypothèses sur les *issues intéressantes* de l'expérience globale.

La compréhension de la juste portée des *hypothèses de modélisation* est un enjeu important dans l'introduction de la notion d'espace probabilisé, c'est à dire, dans les « programmes 2000 », la définition d'un ensemble fini d'issues et la détermination de la probabilité de chacune de ces issues.
Cela amène l'enseignant, pour une même expérience, à envisager avec les élèves plusieurs univers, et à expliquer pourquoi choisir tel univers plutôt que tel autre.

Les premiers exemples abordés sont des exemples de lancers de plusieurs dés ou de plusieurs pièces, les objets lancés étant supposés équilibrés, ou des exemples de plusieurs tirages successifs d'un jeton d'une urne avec remise, chaque tirage se faisant au hasard.

Les hypothèses de modélisation amènent à choisir l'univers des *issues physiques* **parce que la loi sera équirépartie sur cet univers.**
Cette équiprobabilité des *issues physiques* est présentée comme « de bon sens » : il n'y a pas plus de chances que sorte 2 sur le dé bleu et 3 sur le dé rouge que sorte le double 6, etc.

L'équiprobabilité associée aux issues de chacune des deux expériences élémentaires « lancer d'un dé équilibré » et « lancer du deuxième dé équilibré », fait partie des connaissances spontanées des élèves, mais le passage des *hypothèses de modélisation* qui portent sur chacun des deux dés au résultat « les *issues physiques* de l'expérience globale sont équiprobables » passe par la définition de l'espace probabilisé produit à associer à la composée d'expériences aléatoires indépendantes.
Or la notion d'espace produit ne fait pas partie des programmes : on ne peut alors que faire appel « au bon sens » des élèves pour assurer ce passage de la loi équirépartie pour les espaces probabilisés associé au lancer de chacun des dés, à la loi équirépartie pour l'espace probabilisé produit.

9.3. Le « transport » d'une loi de probabilité par une variable aléatoire

Dans l'exemple du lancer de deux dés associé à divers jeux, il est possible de travailler avec pour univers l'ensemble des *issues intéressantes*, après avoir déterminé la loi de probabilité sur les *issues physiques* : on définit alors une variable aléatoire qui à chaque élément de l'univers (ici l'ensemble des *issues physiques*), associe un élément (qui caractérise un *événement intéressant*), l'ensemble des images sera le nouvel univers (l'univers des *issues intéressantes*).
Pour le jeu du « Monopoly », la variable aléatoire à considérer est la variable S qui au couple (x ; y) de Ω associe x + y ; l'*événement intéressant* : « avoir une somme égale à 7 » était représenté par l'ensemble {(1 ;6) ; (6 ; 1) ; (2 ; 5) ; (5 ; 2) ; (3 ; 4) ; (4 ; 3)}, de probabilité 6/36 ; cet *événement intéressant* devient une *issue intéressante* pour l'univers Ω' : {2 ; 3 ; … ; 12}, de probabilité 6/36.

Les *hypothèses de modélisation* portent directement sur Ω si l'on se place dans la théorie des probabilités où $\Omega = \{1 ; 2 ; \ldots ; 6\} \times \{1 ; 2 ; \ldots ; 6\}$ et $p = p_1 \otimes p_2$ où p_i est la loi définie sur $\{1 ; 2 ; \ldots ; 6\}$ par le choix des *hypothèses de modélisation* sur chacun des deux dés (et plus généralement pour chacune des expériences élémentaires composant l'expérience aléatoire) ;
Le produit des probabilités p_1 et p_2 est défini par $p_1 \otimes p_2 (x ; y) = p_1(x) \times p_2(y)$, ce qui nous donne bien 1/36 pour chacune des issues de l'univers Ω. On constate alors que la loi est équirépartie sur l'univers Ω à 36 éléments.
Quand on ne dispose pas de ces éléments de la théorie des probabilités (les espaces probabilisés produits), les *hypothèses de modélisation* ne portent plus directement sur Ω au

sens de « sont impliquées par un calcul mathématique », et l'équiprobabilité des *issues physiques* de l'expérience globale induite par les *hypothèses de modélisation* va pouvoir poser un problème aux élèves, que l'enseignant tentera de résoudre « en mettant de la couleur » sur les dés, et en faisant appel à des connaissances spontanées des élèves (le « bon sens »), connaissances spontanées réelles ou supposées.

9.4. La question du choix de l'*univers des possibilités* chez Alarcon

Analysant plusieurs situations imaginées pour étudier le jugement probabiliste de jeunes, Alarcon (1982) détaille les difficultés, dans les situations de prévision, pour choisir l'univers des possibilités, difficultés qu'il estime moindres dans les situations de décision.

• Les situations de prévision correspondent à la situation suivante : deux sacs contiennent des boules blanches et des boules noires ; la composition des sacs est connue ; on demande de choisir parmi les deux sacs celui dans lequel il est le plus probable de tirer une boule blanche. Alarcon présente trois univers de possibilités, enchevêtrés :

- la possibilité de faire un choix ou de ne pas faire de choix,
- les deux sacs (c'est vers cet univers qu'oriente la question, dit Alarcon),
- les événements « tirer une boule blanche » et « tirer une boule noire » (le résultat du tirage oriente vers cet univers).

• Les situations de décision décrites par Alarcon consistent à effectuer un, ou plusieurs tirages, de l'un des sacs (la composition d'au moins un sac est connue), de donner le résultat de ces tirages, et de retrouver, parmi les deux sacs, celui dans lequel a été fait le ou les tirages. Alors dit Alarcon : « *l'incertitude concerne uniquement le choix du sac pris lors du tirage et non plus le résultat du tirage (autrement dit, on n'est plus amené à accepter l'incertitude d'un tirage au hasard)* ».

Le problème du choix de l'univers lorsqu'on étudie le lancer de deux dés et que l'on s'intéresse à la somme des points amenés ne relève pas tout à fait de la même problématique.

• Il ne s'agit ni d'une situation de prévision, ni d'une situation de décision : il est demandé à l'élève de déterminer la probabilité de 11 événements : avoir une somme égale à 2, avoir une somme égale à 3, etc., autrement dit de déterminer la loi de probabilité à associer à une expérience.

• Une seule expérience aléatoire est en jeu : le lancer de deux dés, et l'objectif est de construire un univers des possibilités, alors que dans les situations décrites par Alarcon, il y a deux expériences aléatoires, et un univers associé à chacune de ces expériences.

• L'enjeu, avec la somme des points amenés en lançant deux dés, est de choisir l'univers qui va permettre de traduire les *hypothèses de modélisation* faites :

- si ces hypothèses de modélisation portent sur chacun des deux dés, les univers Ω_1 et Ω_2 à associer à chacun des deux dés sont déterminés, et l'univers à associer à l'expérience s'impose : $\Omega = \Omega_1 \times \Omega_2$, et la probabilité p est définie par $p = p_1 \otimes p_2$;
- mais les hypothèses de modélisation peuvent porter sur les valeurs prises par la somme (par exemple dans le cas où les dés sont pipés, et où l'on a mené une étude statistique des valeurs prises par la somme des points amenés, la probabilité de chacune de ces valeurs est estimée par la fréquence d'apparition de cette valeur).

Dans le premier cas, il est nécessaire de « remonter » aux *issues physiques*, avec la difficulté du contournement de la notion des espaces probabilisés produits (qui ne figure pas dans les programmes) pour déterminer la loi de probabilité des *issues intéressantes* ; dans le second cas, le détour par les *issues physiques* non seulement ne s'impose pas, mais n'est pas possible.

9.5. Imitations

Nous appellerons imitation d'une expérience aléatoire une expérience aléatoire ayant même apparence que l'expérience d'origine (en général : mêmes *issues intéressantes*), sans que soient envisagées des hypothèses de modélisation à associer à l'expérience, ou sans que soit mené un calcul sur les chances de ces issues, ou encore avec une évaluation incorrecte de ces chances.

Ce que nous définissons ainsi est donc une expérience aléatoire considérée uniquement du point de vue de ses issues.

Une imitation pourra être considérée comme une simulation de l'expérience aléatoire étudiée si et seulement si elle produit des résultats, et est une expérience aléatoire équivalente à l'expérience aléatoire initiale (cf. la définition du paragraphe 5).

Par exemple, tirer au hasard un papier parmi 11 papiers numérotés de 2 à 12 est une imitation du lancer de deux dés lorsqu'on s'intéresse à la somme des points amenés, mais n'en est pas une simulation.

Dans l'exemple de la simulation du « jeu du lièvre et de la tortue », les algorithmes proposés ont été construits avec comme fil directeur une imitation de la manière la plus fidèle possible de chacun des « moments » du jeu. Nous aurons à préciser cette technique de simulation.

Nous dirons que nous avons construit une imitation au plus près (ou au plus proche) de l'expérience aléatoire. L'expression « *procédure aussi "proche" que possible de celle* » de l'expérience aléatoire ou l'expression « *procédures de simulation qui soient **aussi "isomorphes" que possible** à l'expérience simulée* » sont employées par Parszysc (2009) dans l'article que nous avons analysé au paragraphe 3 du chapitre 4.

10. Pré-modélisation d'une expérience

Nous appellerons *pré-modélisation* de l'expérience aléatoire, l'étape où sont émises les hypothèses de modélisation.

C'est une étape préliminaire incontournable lorsque l'on veut modéliser ou simuler cette expérience aléatoire.

- Il s'agit d'abord de décomposer l'expérience en expériences élémentaires.
- Il faut ensuite émettre des hypothèses permettant en quelque sorte de mathématiser (ou au moins coder) les issues possibles de chacune des expériences aléatoires élémentaires qui ont été dégagées (par exemple, la pièce ne peut pas tomber sur la tranche, ou bien le dé ne peut pas retomber sur une arête).

Il faut également interpréter certains éléments du dispositif expérimental, tels que : lorsqu'on lance simultanément deux dés, on peut négliger l'effet des chocs et considérer que « les faces sorties sont indépendantes », ou bien encore : si on répète plusieurs fois le lancer d'un dé en agitant le gobelet contenant le dé avant chaque lancer, on peut supposer que les lancers réalisent des expériences deux à deux indépendantes, etc..

- Il s'agit ensuite de choisir les hypothèses de modélisation pour chacune de ces expériences élémentaires : les dés sont équilibrés et la loi associée à chacun est équirépartie, ou au contraire, l'un des dés est truquée, et par exemple, la loi qui lui est associée est (1/7 ; 1/7 ; 1/7 ; 1/7 ; 1/7 ; 2/7).

Cette étape de *pré-modélisation* consiste donc essentiellement à déterminer les expériences aléatoires élémentaires qui composent l'expérience aléatoire à modéliser ou à simuler, et à faire sur ces expériences élémentaires les hypothèses qui permettront de leur associer une loi de probabilité.

Cette étape de pré-modélisation est commune, et nécessaire aux deux activités : modélisation et simulation.

CHAPITRE 8

Le couple modélisation / simulation

> *« Formellement, simuler une expérience, c'est choisir un modèle pour cette expérience – ici une loi de probabilité P sur un certain ensemble – puis produire une liste de données pour lesquelles P est un modèle pertinent (...) »*
> **Document d'accompagnement du programme de Seconde publié dans le B.O. hors-série, n°6, du 12 août 1999.**

> « *(...) La simulation remplaçant l'expérimentation permet, avec une grande économie de moyens, d'observer des résultats associés à la réalisation d'un très grand nombre d'expériences. (...)* »
> **Programme de Seconde**

1. L'analyse de Coutinho

Au début de sa thèse, Coutinho (2001) analyse les liens entre modèle probabiliste et simulation d'une expérience aléatoire. Elle s'inscrit dans la définition donnée par les « programmes 2000 », citée ci-dessus. Par conséquent une « simulation informatique » peut être considérée comme une expérience aléatoire, mais sa qualité de *simulation* d'une expérience aléatoire ne peut être que « de principe » :

> *L'environnement informatique (calculatrice ou ordinateur) permet de reproduire une expérimentation facilement, un grand nombre de fois et dans des conditions d'indépendance* [55].
> *Mais dans la situation où les paramètres de cette expérience ne sont pas donnés ou ne sont pas accessibles, n'étant pas connus, la simulation évoquée ne peut être que de principe. Puisqu'ils ne sont pas connus, ils ne sont pas injectés dans le*

[55] Dans la mesure où Coutinho émet des réserves sur l'aléatoire ou le pseudo aléatoire produit par un ordinateur, cette indépendance des essais n'est pas acquise d'emblée lorsque les nombres pseudo-aléatoires sont générés par une suite récurrente.Pour notre part, nous considérons une suite de nombres pseudo-aléatoires selon une loi équirépartie comme simulant une suite de nombres aléatoires équirépartis d'une manière acceptable dans l'enseignement secondaire, le calcul de deux nombres pseudo-aléatoires étant associé à deux expériences aléatoires indépendantes, même si ces deux nombres sont deux valeurs d'une suite définie par récurrence.

fonctionnement de l'artefact informatique. »

Lorsque Coutinho écrit sa thèse, il n'était pas question, dans la lettre des programmes auxquels elle se réfère, que la tâche de simulation soit dévolue à l'élève. Coutinho a conçu dans sa thèse une ingénierie didactique utilisant des simulations informatiques qui sont des logiciels « fermés », au sens où l'élève n'a pas accès au programme qui génère les simulations : il peut simplement intervenir sur les valeurs de certains paramètres de la simulation, paramètres qui sont des données pour l'exécution du programme (nombre de répétitions de la simulation par exemple). C'est le même type de logiciels-simulations qu'évoquent dans leurs thèses Zaki 1990 et Bordier (1991).

Citons à nouveau Coutinho :

> *Dans ce cas, nous ne parlerons pas de simulation de l'expérience car il n'y a pas de modèle inséré dans cet environnement : il s'agit de la production d'une représentation seulement évocatrice de l'expérience à l'écran. En réalité, nous sommes en train de réaliser une expérience aléatoire intrinsèque, en la reproduisant un grand nombre de fois grâce à la performance de l'artefact informatique choisi.*

La question est à nouveau posée : à quel moment une *imitation* d'une expérience aléatoire obtient-elle le statut de *simulation* d'une expérience aléatoire ?
Coutinho semble considérer que tant que la preuve n'est pas administrée que cette *imitation* a même modèle que l'expérience aléatoire, elle ne peut en être une *simulation* . Citons encore Coutinho :

> « *Donnons un exemple. Avec une calculatrice, nous pouvons utiliser la touche « rand » pour obtenir des valeurs entières entre 1 et 6. Supposant que nous n'avons pas accès aux formules installées pour la fabrication des nombres pseudo-aléatoires qui seront affichés à l'écran lors de l'utilisation de cette touche, nous considérons cette manipulation de la calculatrice comme une expérience aléatoire en soi. Nous ne cherchons pas nécessairement à la concevoir comme la simulation d'un jeu de dés.*
>
> *Pour utiliser la calculatrice comme outil de simulation de ce jeu, il faudrait :*
>
> *• concevoir le jeu avec le dé comme une expérimentation réelle ;*

• ensuite, construire des hypothèses de modèle sur l'équiprobabilité des faces possibles, déjà dans un processus d'abstraction ;

• s'engager dans le processus de construction d'un modèle théorique pour ce jeu, modèle qui sera injecté dans le mode d'emploi de la touche « rand » pour aboutir aux six chiffres équirépartis. *».*

Dans cette optique, de telles « simulations » sont considérées comme des expériences aléatoires d'utilisation facile, qui peuvent se substituer à l'expérimentation réelle pour illustrer des notions du programme telle que le notion de fluctuation d'échantillonnage. On peut en quelque sorte oublier l'expérience d'origine simulée, en s'abstenant d'interpréter l'expérience informatique comme une *simulation* de cette expérience :

> *Remarquons qu'il ne s'agit pas de modifier l'action effective réalisée sur l'ordinateur ou sur la calculatrice : on change le point de vue lorsqu'on change l'enjeu de cette action par le changement de son interprétation. On passe d'une simulation pour laquelle le modèle théorique de l'expérimentation est déjà prêt, à la reproduction d'une expérimentation dont l'objectif est de déclencher une démarche de modélisation par l'observation de ses invariants.*
>
> *Il n'existe pas de véritable simulation sans un processus de modélisation préalable.*
>
> *Le changement de point de vue exposé ci-dessus, qui préconise d'expliciter la distinction entre simulation et production d'une expérience informatique, est conforme aux objectifs du programme : faire comprendre à chaque élève la signification de l'action de simuler une expérience de la réalité. Citons un extrait du document d'accompagnement du programme qui explique bien le statut de la simulation (...)* » Coutinho (2001)
>
> L'extrait cité est l'extrait que nous avons placé au début de ce chapitre.

Nous avons souligné que le point de vue de Coutinho s'inscrit dans un contexte où la simulation n'est pas conçue par l'élève, qui se trouve donc en position d'utilisateur d'un logiciel qui lui donne à voir à l'écran une imitation de l'expérience aléatoire ; l'élève reconnaîtra les issues intéressantes de l'expérience aléatoire qu'il veut étudier, mais il ne pourra pas comparer la loi de probabilité associée à l'expérience générée par l'ordinateur, à l'expérience aléatoire qu'il veut étudier (il ne peut qu'observer les résultats générés par la simulation).

Dans ce contexte, il est de fait impossible de décider si une expérience informatique est une simulation d'une expérience aléatoire donnée ; éventuellement une confrontation, par le test du χ^2 par exemple, des données produites par « l'artefact informatique », comme dit Coutinho, avec la loi de probabilité pourrait, à un risque donné, valider l'hypothèse que l'artefact est bien une simulation de l'expérience aléatoire ; mais cette démarche nécessite que l'on connaisse la loi de probabilité à associer à l'expérience aléatoire.

Si comme semble l'affirmer Coutinho et l'extrait du document d'accompagnement cités ci-dessus, la détermination du modèle probabiliste est un préalable à la création de toute simulation (informatique ou non) d'une expérience aléatoire, alors, comme nous l'avons déjà évoqué, et comme l'indique l'extrait du programme de seconde placé au début de ce chapitre, les simulations n'ont leur place dans l'enseignement secondaire que comme substitut à une expérimentation laborieuse.

Cependant le fait que les simulations informatiques passent du statut de « logiciel fermé » au statut de « logiciel ouvert » lorsque c'est l'élève qui construit le simulation, permet de poser la question de la « qualité de simulation » dans d'autres termes.
Nous avons vu en effet, dans le chapitre 6, l'élaboration d'une simulation du « jeu du lièvre et de la tortue » qui n'était pas fondée sur la détermination préalable de la loi de probabilité à associer à ce jeu.
Nous avons pu démontrer que l'expérience informatique que nous avons générée a même loi de probabilité que l'expérience du « jeu du lièvre et de la tortue », ce qui a confirmé que cette expérience informatique est bien une simulation de l'expérience du « jeu du lièvre et de la tortue ».
Nous avons pu faire cette démonstration parce que le « logiciel » ainsi produit était ouvert : ayant accès au **programme** dont l'exécution génère une expérience aléatoire, nous avons pu déterminer la loi de probabilité de cette expérience aléatoire, et la comparer à la loi de probabilité associée à l'expérience d'origine.

Cependant, nous postulons que l'expérience aléatoire générée par ce programme est une simulation de l'expérience aléatoire d'origine <u>par construction</u> de ce programme, et plus précisément par <u>construction de l'algorithme</u> que ce programme traduit dans un langage de programmation donné.
Pour éclairer cette affirmation nous devons essayer de préciser la technique qui permet de

construire un tel algorithme.

2. *Imitation au plus près*

2.1. la technique mise en œuvre

Nous avons décrit au paragraphe 9.5. du chapitre précédent cette technique comme une technique qui produit une *imitation au plus près* de l'expérience aléatoire à simuler.
Dans l'étape de pré-modélisation, on repère les expériences aléatoires élémentaires qui interviennent dans l'expérience globale, si celle-ci est une expérience composée.

L'expérience est constituée d'un enchaînement, avec un ordre chronologique, d'un certain nombre d'actions, certaines ayant un résultat aléatoire (ce sont les expériences aléatoires élémentaires repérées lors de la pré-modélisation) : dans le « jeu du lièvre et de la tortue », on lance le dé une première fois ; si sort un six, on déclare le lièvre gagnant, sinon, on fait avancer la tortue d'un pas, puis on relance le dé etc.

Le résultat de la tâche de simulation réalisée selon la technique que nous voulons décrire, doit « imiter au plus près » chacune de ces actions et les liens entre ces actions, en rendant correctement compte de leur ordre chronologique, qui doit aussi être « imité au plus près ». Cela exige que l'expérience aléatoire à simuler ait été décrite de façon suffisamment détaillée pour que l'on puisse déduire de cette description toutes ces actions et leurs liens, y compris les enchaînements chronologiques.

Cette description, nous pouvons l'appeler *le cahier des charges* associé à l'expérience aléatoire, dans la mesure où cette description est ce que celui qui écrira le programme, le programmeur, devra traduire en un algorithme, algorithme traduit ensuite en un programme, par le programmeur toujours, dans un langage de programmation donné.

Nous avons repris l'expression qui a cours dans le « monde informatique » . Citons Morand (2009) :

> *« Un programme ou un ensemble de programmes constitue une réponse, partiellement ou totalement automatisée, aux besoins d'un groupe social identifiable au sein d'une collectivité plus ou moins vaste. (…) »*

L'auteur prend l'exemple du dossier médical informatisé

> *« Le développeur de logiciels n'a pas en général les compétences associées au métier de médecin, de personnel soignant, de financier ou de responsable d'une politique de santé. Il agit donc nécessairement comme intermédiaire entre des demandes exprimées par les utilisateurs du futur programme (supposées cohérentes) et les fonctionnalités que ce logiciel offrira.*
>
> *(…)*
>
> *Dans les débuts de l'informatique, on a souvent pensé que la partition métier demandeur/spécialiste informatique pouvait s'institutionnaliser en deux entités distinctes, parties prenantes d'un même contrat : « le cahier des charges ». On a ainsi pensé pouvoir fixer une répartition des rôles dans laquelle le demandeur est responsable du contenu de la demande et l'informaticien est en charge de sa réalisation technique. »*

Ce qui nous intéresse ici, est la mise en évidence de l'existence d'un « écart » entre une demande : simuler une expérience aléatoire, et le programme dont l'exécution sera une simulation de l'expérience. Entre la demande et le programme, il y a cette interface que constitue le *cahier des charges* ; le « développeur » aura pour tâche d'en assurer « la réalisation technique », au travers de l'élaboration d'un algorithme qu'il traduira, dans un langage de programmation, en un programme exécutable par la machine.

La demande ici va être une demande de l'institution scolaire à un élève ou un groupe d'élèves, ou au professeur qui seront les programmeurs (les « développeurs »). L'institution scolaire « demandeur » pourra être incarnée par le professeur, par la classe, par l'élève ou un groupe d'élève, ou par un énoncé d'examen.

La décomposition de la tâche par cette technique de simulation consiste en :
(1) la définition de la demande
(2) l'analyse du problème afin de lui associer un *cahier des charges*,
(3) l'élaboration d'un algorithme, étape intermédiaire entre le *cahier des charges* et le programme,
(4) la traduction de cet algorithme en un programme exécutable.

Cette présentation est très schématique ; en fait, ces étapes vont éventuellement s'enchevêtrer, d'autant plus que le « demandeur » peut être également le programmeur, et que les algorithmes à élaborer pour les simulations envisageables au lycée sont très simples. De plus, en général la tâche s'inscrit dans des contextes particuliers, dans le cadre de l'institution scolaire. Précisons par des exemples :

• Lorsque l'expérience aléatoire est simple, la distinction entre la demande (« simulez cette expérience aléatoire »), le *cahier des charges*, et l'algorithme peut devenir complètement artificielle ; le cas extrême est la simulation du lancer d'un dé cubique équilibré par l'appel « Alea.entre.bornes(1 ; 6) ». Rappelons que nous prenons comme hypothèse que les nombres pseudo-aléatoires calculés par un ordinateur sont « acceptables », au sens où nous considérons que l'appel « Alea.entre.bornes(n ; p) où n et p sont deux entiers, rend un nombre entier aléatoire entre n et p, selon la loi équirépartie.

• Lorsque le professeur, ou les « documents d'accompagnement » proposés aux enseignants, souhaitent simuler une expérience aléatoire pour illustrer la loi des grands nombres, le demandeur est aussi le « développeur » et a à sa charge les quatre étapes.

• Lorsque le professeur demande à ses élèves, en devoir à la maison, de simuler le « jeu du lièvre et de la tortue », la deuxième étape, qui devrait être à la charge du « demandeur », va être reportée à la charge du programmeur : elle fait partie de l'exercice que le professeur propose à l'élève.

• Lorsque la classe a à résoudre un problème de prise de décision dans une situation d'incertitude où intervient une expérience aléatoire, telle que « combien faut-il acheter de tablettes de chocolat pour avoir la collection d'images complète ? », un groupe peut décider de simuler l'expérience aléatoire qui consiste à acheter des tablettes de chocolat tant que la collection d'image n'est pas complète. Ce groupe se trouve alors prendre en charge les quatre étapes.

Morand (2009) déjà cité, souligne que la répartition des rôles entre le « demandeur » et le « développeur » est en fait plus complexe :

> *« (...) le demandeur ne sait pas ce qu'il est en droit d'attendre de l'informaticien et ce dernier se trouve bien en peine de comprendre la nature des problèmes pour lesquels il a été sollicité. »*

Ce constat fait qu'il y a nécessité d'une collaboration entre le « demandeur » et le programmeur, qui ne s'arrête pas à la donnée par le premier au second d'un morceau de papier sur lequel serait écrit ce *cahier des charges* : l'élaboration de ce *cahier des charges* est

une œuvre menée conjointement par le « demandeur » et le « développeur ».
Toutes proportions gardées dans cette analogie entre le monde professionnel et le contexte de la classe, c'est ce qui va se produire lorsque la simulation de l'expérience aléatoire n'est pas un exercice « en soi », mais a un objectif : l'illustration de la loi des grands nombres par exemple ou plus généralement la production de données empiriques qui pourront être exploitées, par exemple par une étude statistique.
Dans ce cas, il va falloir adjoindre d'autres demandes au *cahier des charges* associé à l'expérience aléatoire, en général au minimum la répétition de la simulation, assortie d'un traitement des résultats ainsi générés.

Ces demandes peuvent être prévues à l'avance, ou décidées « après coup », ou complétées au fur et à mesure : « nous avons réussi à simuler la simulation de l'achat des tablettes de chocolat jusqu'à obtention de la collection complète, comment exploiter cette simulation ? »
- Ce peut être en calculant le nombre moyen d'achats nécessaires pour avoir la collection complète, ce qui suppose que l'algorithme programmé contienne le calcul, pour chaque simulation, du nombre de tablettes achetée ; le calcul de la valeur moyenne peut se faire sans consommer d'espace mémoire. Il y a un rajout au *cahier des charges* qui se traite pratiquement immédiatement.
- Ce peut être en décidant de mener une étude statistique plus fine. Les indicateurs statistiques devront alors être définis, ainsi que les conditions de calcul de ces indicateurs : les données empiriques fruits de la simulation répétée seront-elles transférées à un module spécifique de calcul statistique (utilisation d'une bibliothèque de programmes), ou les calculs statistiques doivent-ils être inclus dans le programme de la simulation ? Ces questions relèvent du *cahier des charges*. Elles n'y sont peut être pas inscrites d'emblée.

Remarquons que le *cahier des charges* n'inclut pas nécessairement les *hypothèses de modélisation* à associer à chacune des expériences aléatoires élémentaires.
• Ces hypothèses peuvent être énoncées dans le *cahier des charges*, à charge pour le programmeur de les « traduire » correctement (le *cahier des charges* spécifie : « on lance deux dés équilibrés ») ;
• ces hypothèses peuvent être à la charge du programmeur (le *cahier des charges* spécifie : « on lance deux dés »), le programmeur décidera des *hypothèses de modélisation* qu'il choisit pour *l'imitation* de chacune des expériences aléatoires élémentaires ;
• ces hypothèses peuvent enfin être à la charge de celui qui fait exécuter le programme : cela

doit alors être spécifié dans le *cahier des charges* : « on lance deux dés, (…) ; l'utilisateur de la simulation pourra décider des lois de probabilité à associer à chacun des dés » ; le programmeur aura alors prévu ces *hypothèses de modélisation* comme paramètres de son programme.

Mais là encore, la répartition des rôles peut être ici bien artificielle, surtout si c'est la même personne qui assume la plupart des étapes de cette technique.

2.2. La simulation obtenue

Il est important de souligner ici à nouveau qu'il y a dans toute simulation d'une expérience aléatoire, comme dans la détermination d'un modèle probabiliste de cette expérience un choix nécessaire, celui des *hypothèses de modélisation*. Ce choix se fait dans l'étape de pré-modélisation qui est une étape commune aux deux tâches de modélisation probabiliste et de simulation.
Cette étape peut être menée une seule fois, en particulier lorsque la technique de simulation employée est la simulation de la loi de probabilité.
Elle peut être menée deux fois, une fois pour chaque tâche ; les mêmes choix d'hypothèses devraient être faits pour l'une et l'autre des tâches. C'est ce qu'il pourra être intéressant d'observer lorsqu'un élève doit simuler et modéliser une même expérience aléatoire.

Bien entendu, une *imitation au plus près* d'une expérience aléatoire ne pourra prétendre avoir même modèle probabiliste que cette expérience aléatoire que lorsque les mêmes *hypothèses de modélisation* auront été faites pour déterminer et la loi de probabilité, et la simulation de cette expérience aléatoire.

Implicitement, dans ce qui précède et ce qui va suivre, lorsque nous rapprochons, pour une même expérience aléatoire, sa simulation et sa loi de probabilité, les mêmes hypothèses de modélisation ont été faites pour déterminer l'une et l'autre, que, dans la pratique, ces hypothèses aient été posées une fois pour les deux tâches, ou qu'elles aient été posées par deux fois.

Nous obtenons, par la technique qui vient d'être décrite au paragraphe précédent, un programme dont l'exécution est une expérience aléatoire pour laquelle <u>nous postulons</u> que

c'est une simulation informatique parmi d'autres simulations (informatiques ou non) possibles sous les mêmes *hypothèses de modélisation*, en particulier la simulation qu'on obtient en simulant la loi de probabilité de cette expérience.
Cette dernière simulation est, comme nous l'avons fait remarquer au chapitre 7, une *simulation* de l'expérience aléatoire initiale en quelque sorte par définition d'une simulation : c'est une expérience aléatoire qui a même loi de probabilité que l'expérience simulée.

Une *simulation au plus près* d'une expérience aléatoire, en est bien une simulation, c'est à dire une expérience aléatoire ayant même modèle probabiliste, mais par construction. Autrement dit nous postulons que c'est la technique même de construction de cette *imitation au plus près* qui va assurer que l'expérience aléatoire ainsi obtenue a bien même modèle que l'expérience aléatoire d'origine, alors que le lien qui fait que l'imitation est bien une simulation, c'est à dire la loi de probabilité de l'expérience, n'a peut-être même pas été déterminé.

Cette affirmation demande bien sûr à être étayée par un discours d'ordre technologique sur la technique employée. Nous tenterons de nous y employer dans la quatrième partie.

On pourrait admettre que cette affirmation conserve son statut de *postulat*. On pourrait dire que dans le cas où la simulation se fait par une *imitation au plus près* de l'expérience aléatoire, le lien entre la simulation et le modèle probabiliste **est de principe.**
C'est ce que font implicitement les documents pour le professeur qui accompagnent les « nouveaux programmes », les « programmes 2010 » : cf [1] et [2].
Les « programmes 2010 » reprennent certes la définition des « programmes 2000 » : simuler une expérience aléatoire, c'est simuler sa loi de probabilité, cependant les exemples de simulations informatiques qui sont proposés dans les documents qui accompagnent ces nouveaux programmes sont *des imitations au plus près* de l'expérience aléatoire,

- sans que soit jamais évoquée la technique qui mène à ces simulations informatiques,
- et sans que soit non plus jamais évoquée la question de la **preuve** que les expériences aléatoires ainsi générées ont même loi de probabilité que l'expérience d'origine.

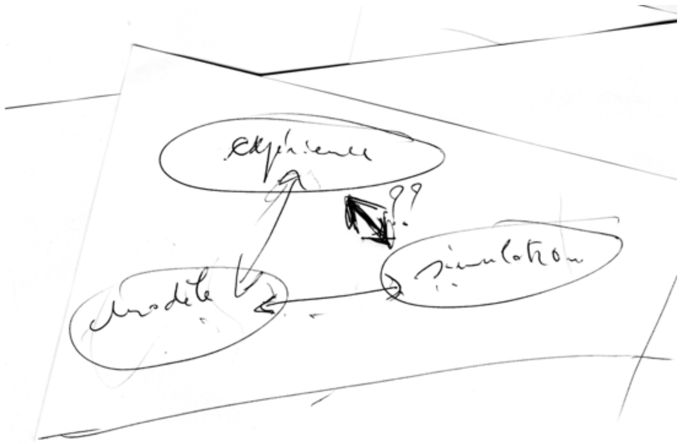
Nous présenterons ces simulations proposées par les documents d'accompagnement au chapitre 11, et verrons qu'il y a une autre absence dans ces documents : une réflexion sur les problèmes didactiques posés par l'introduction de l'algorithmique-programmation, même envisagée de façon « transversale », dans l'enseignement au lycée, en particulier au travers de l'élaboration par les élèves de simulations informatiques d'expériences aléatoires.

En effet, **l'activité de simulation ainsi conçue :** ***imitation au plus près*** **de l'expérience aléatoire, est une activité essentiellement de type algorithmique.**

CHAPITRE 9

Les liens entre les différents champs : réalité / modèle / simulation

Reprenons le questionnement posé par le triangle



Quels sont les liens entre l'expérience et sa simulation ?

1. Un schéma dépassé

Les textes précédemment évoqués (Coutinho (2001), documents d'accompagnement et programmes) suggèrent un lien expérience / simulation « par transitivité » : il y a un lien entre l'expérience aléatoire et le modèle qu'on en propose ; il y a un lien entre ce modèle et la simulation qu'on en fait ; il y a donc un lien entre l'expérience aléatoire et sa simulation, et ce lien est le modèle probabiliste, la loi de probabilité associée à l'expérience, qui serait inscrite dans la simulation de l'expérience aléatoire, comme la loi d'Ohm le serait dans une simulation d'un circuit électrique dans laquelle figurerait la relation U = R I.

Il nous semble que cette présentation a eu sa pertinence lorsque l'ordinateur avait une certaine place dans la classe : une utilisation relativement exceptionnelle, avec un « didacticiel » fermé, tel qu'un logiciel simulant des tirages d'une urne ; l'élève avait alors à paramétrer l'urne (nombre de boules, nombre de couleurs, proportions de boules, etc.). Le passage par une modélisation probabiliste permettant de remplacer l'expérience initiale par un tirage d'urne équivalent était donc un passage obligé.

La situation a complètement changé .
D'une part pratiquement tous les élèves de lycée ont un accès facile à un ordinateur et à une connexion haut débit, en dehors même de l'institution scolaire (ordinateur personnel ou familial, bibliothèques), ce qui leur permet d'utiliser, en dehors de la classe, des logiciels gratuits : logiciels de construction géométrique, langages de programmation (tableur, langages impératifs), logiciels de statistique, etc.
D'autre part les « T.I.C.E. » ont fini par entrer dans la classe de mathématiques : des logiciels sont utilisés pour introduire ou illustrer une notion mathématique, ou pour élaborer des conjectures lors de l'étude d'un problème.
Une maîtrise élémentaire du tableur est un objectif du « B2I » des collèges.

Ces logiciels sont « ouverts » en ce sens que l'élève peut construire lui-même « l'outil » dont il a besoin pour résoudre le problème qu'on lui a soumis, par une activité de type algorithmique : ce peut être la construction de la figure associée à son problème de géométrie ; ce peut être la conception de la feuille de calcul qui calculera les 100 premiers termes d'une suite définie par récurrence dont la convergence est donnée à étudier ; ce peut être le programme de la calculatrice qui simulera le « jeu du lièvre et de la tortue ».

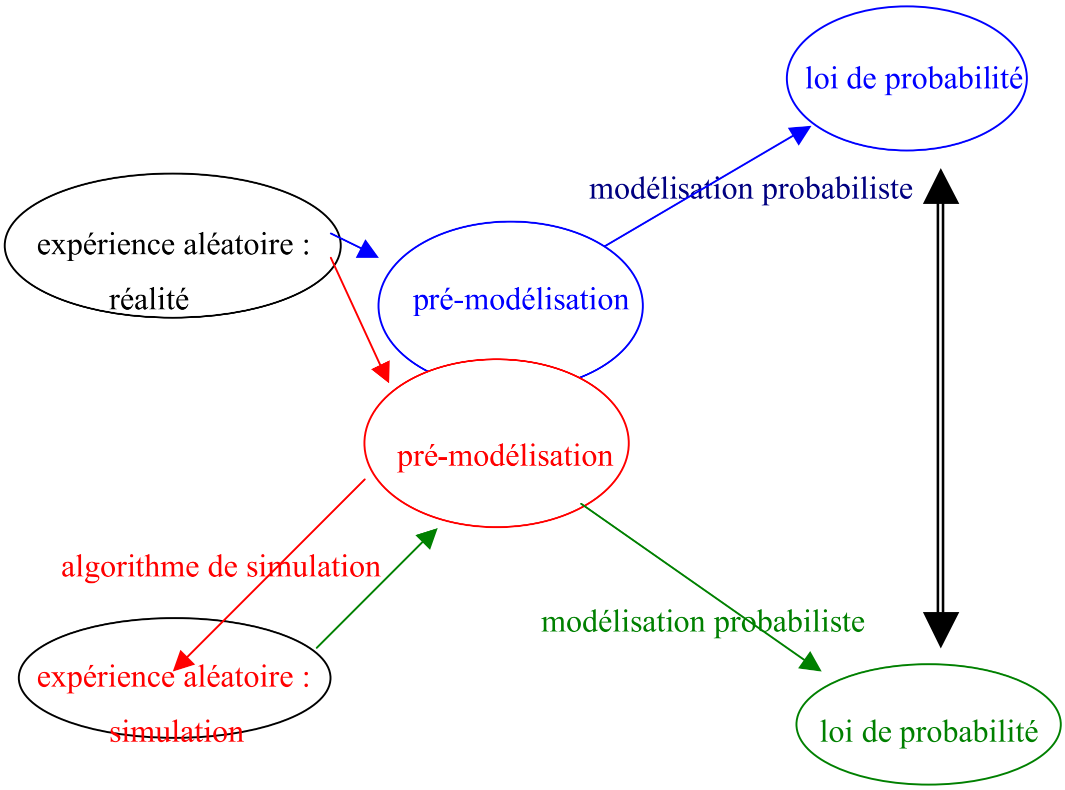
Les programmes de mathématiques de seconde mis en place à la rentrée 2009 ont introduit explicitement des activités d'algorithmique-programmation dans des langages impératifs.
Les élèves de terminale scientifique à la rentrée 2012 auront à choisir une spécialité parmi les quatre spécialités : « mathématiques », « sciences physiques », « sciences de la vie et de la terre », et « informatique et science du numérique » (I.S.N.).
Cette spécialité « I.S.N. » comportera un enseignement d'algorithmique-programmation, enseignement initié en cours de mathématiques en classe de seconde et de première S par des activités « transversales » d'algorithmique-programmation.

Il semble que la question de la simulation d'une expérience aléatoire relève, dans ce contexte, d'un nouveau paradigme : mis à part le choix des *hypothèses de modélisation* associées à chacune des expériences aléatoires élémentaires composant l'expérience globale, l'activité de simulation d'une expérience aléatoire est une activité de type algorithmique qui ne fait pas appel à la théorie des probabilités.

2. La place de l'algorithmique dans une activité de simulation informatique

L'analyse que nous venons de faire de l'activité de simulation telle qu'on peut maintenant l'envisager en lycée, nous amène à proposer les liens du schéma 1 qui suit. Une activité de type algorithmique intervient **es qualité**. Elle ne contient pas a priori le modèle probabiliste de l'expérience aléatoire [56]. L'algorithme à créer doit répondre à un *cahier des charges*.

Schéma 1



• En bleu : l'activité de modélisation probabiliste conduisant à un modèle qui permet de calculer des résultats théoriques : la loi de probabilité de l'expérience de la réalité.

• En rouge : l'activité de création d'une simulation, c'est à dire d'une expérience aléatoire effective, produisant des données empiriques.

• En vert : puisque la simulation est elle-même une expérience aléatoire, on peut déterminer son modèle probabiliste (mais ce n'est pas une nécessité). C'est ce que nous avons fait au chapitre 6, paragraphe 1.6. pour la simulation du « jeu du lièvre et de la tortue ».

[56] même s'il n'est pas interdit de programmer la loi de probabilité déterminée en modélisant, si l'on sait faire, l'expérience. La partie expérimentale de notre thèse a l'ambition de déterminer quel sont les choix des élèves dans une situation « raisonnable ».

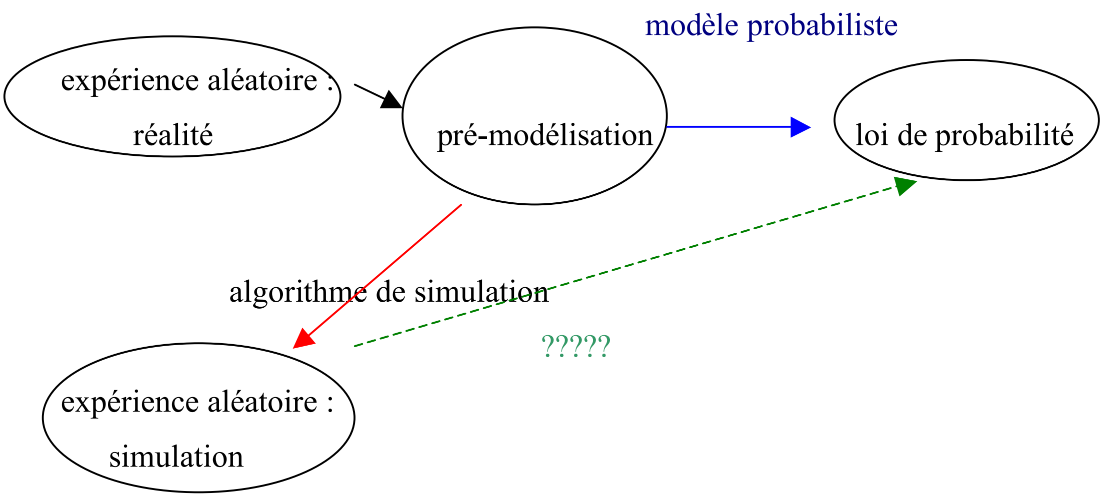
Dans les activités de simulation que nous avons déjà présentées et celles qui nous paraissent « raisonnables » dans l'enseignement secondaire, les deux flèches rouges et bleues partant de la réalité pour aller vers la pré-modélisation pourraient ou devraient être confondues : les pré-modélisations doivent présenter les mêmes réductions [57] et le même choix d'*hypothèses de modélisation* ; les deux expériences aléatoires (l'expérience initiale issue de la réalité et la simulation) auront alors la même loi de probabilité (ce que nous avons marqué par la double flèche noire, mais cette identité ne découle pas de la construction de la simulation comme simulation d'une loi de probabilité préalablement déterminée.

Enfin, la modélisation probabiliste de l'expérience aléatoire « simulation » n'a aucune raison d'être menée.

Les flèches vertes sont alors « de principe ».

On obtient le schéma suivant :

<u>Schéma 2</u>



Nous avons souligné par les points d'interrogation que la preuve que l'expérience aléatoire de la simulation est bien une expérience équivalente à l'expérience initiale pose divers problèmes. On peut, comme indiqué sur le schéma 1, déterminer la loi de probabilité de la simulation et constater que c'est la même que celle de l'expérience de départ. C'est ce que nous avons fait pour le « jeu du lièvre et de la tortue ».

---

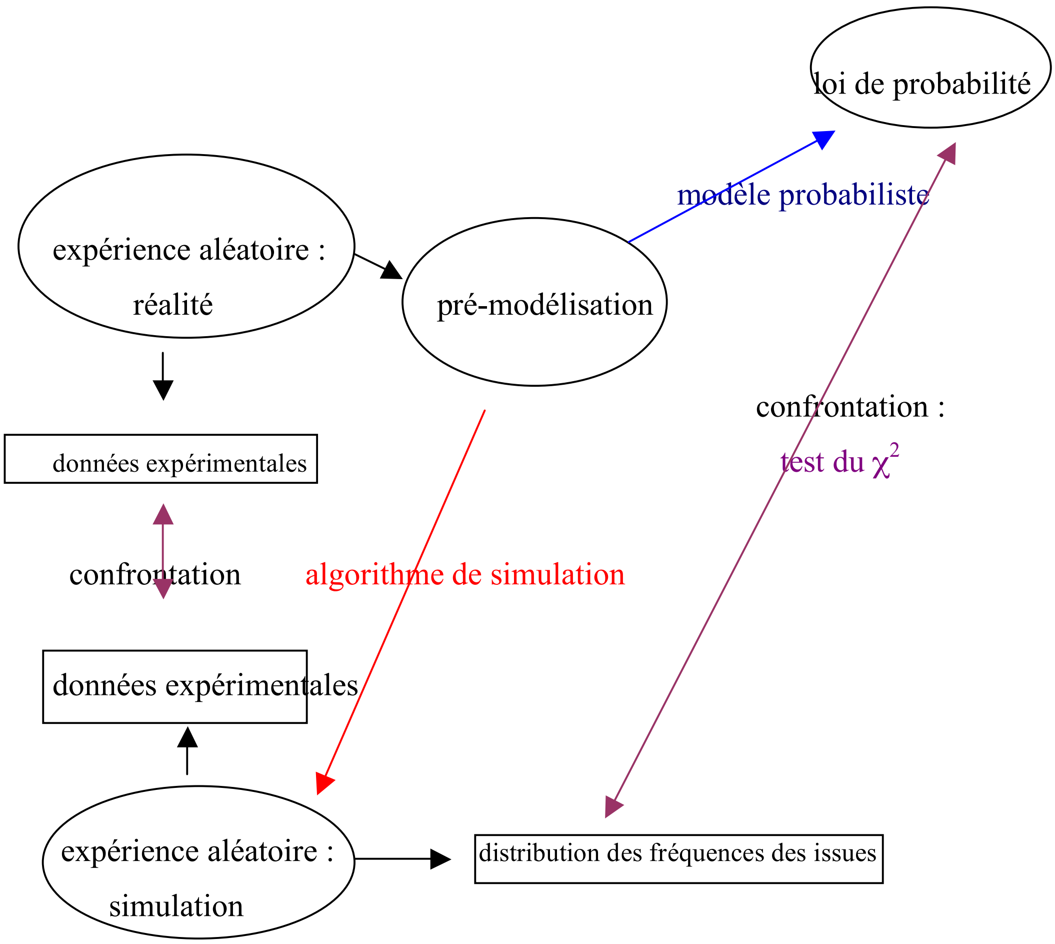
[57] telles que « la pièce ne peut pas retomber sur la tranche » pour le lancer d'une pièce.

On peut également confronter les données empiriques de la simulation

- aux données empiriques de l'expérience de la réalité, s'il y en a
- à la loi de probabilité, si on a pu la déterminer.

Nous illustrons ces confrontations sur le schéma 3.

Schéma 3

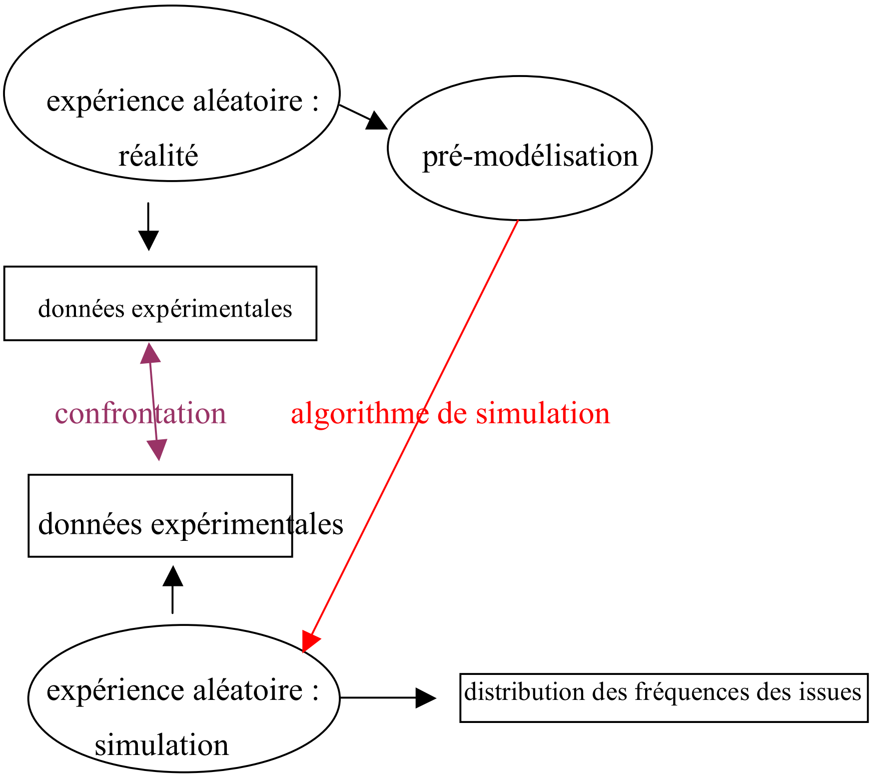


Mais on peut aussi se trouver dans l'un des cas décrits par les schémas 4 et 5 , en l'absence de données expérimentales concernant l'expérience de départ, ou en l'absence de modèle de cette expérience.

Le schéma 6 montre comment on peut utiliser les données empiriques de la simulation pour construire un modèle de l'expérience initiale.

Schéma 4

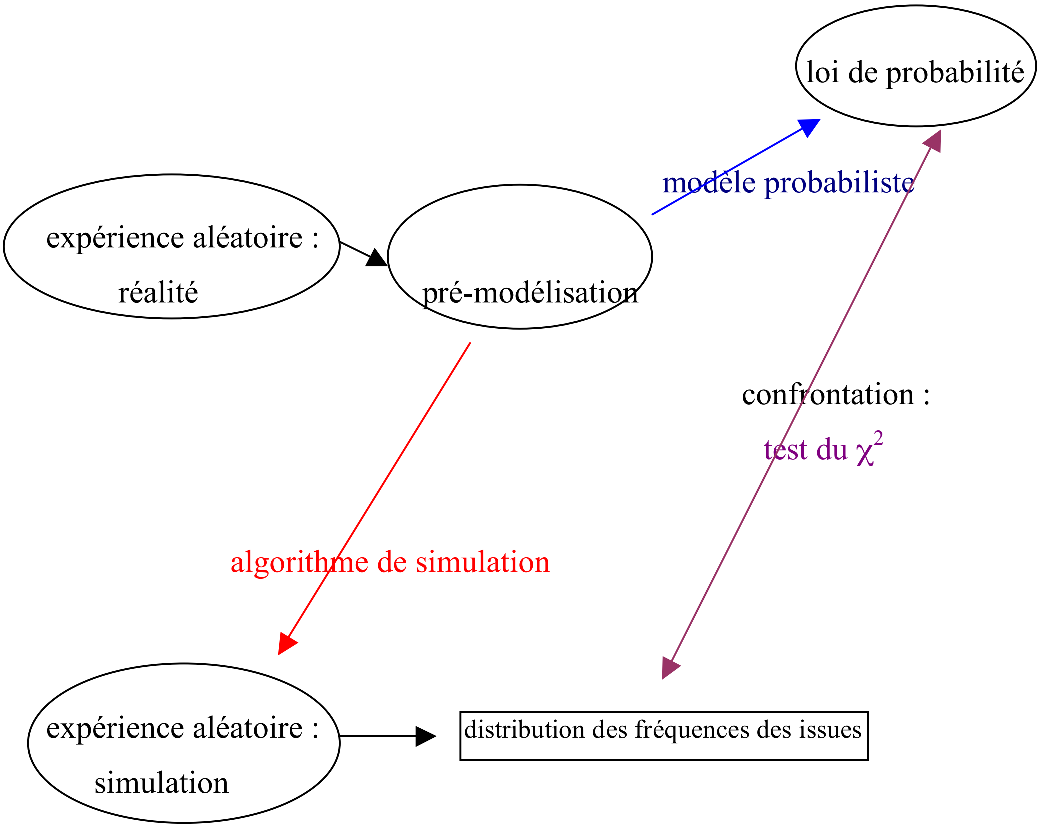
on n'a pas déterminé la loi de probabilité de l'expérience initiale :



La nature de la confrontation entre les données expérimentales des deux expériences n'est pas étudiée ici. On peut imaginer un paramètre calculant la proximité entre ces données. Cela suppose que les ensembles des issues U_1 et U_2 des deux expériences soient correctement mis en bijection.

Schéma 5

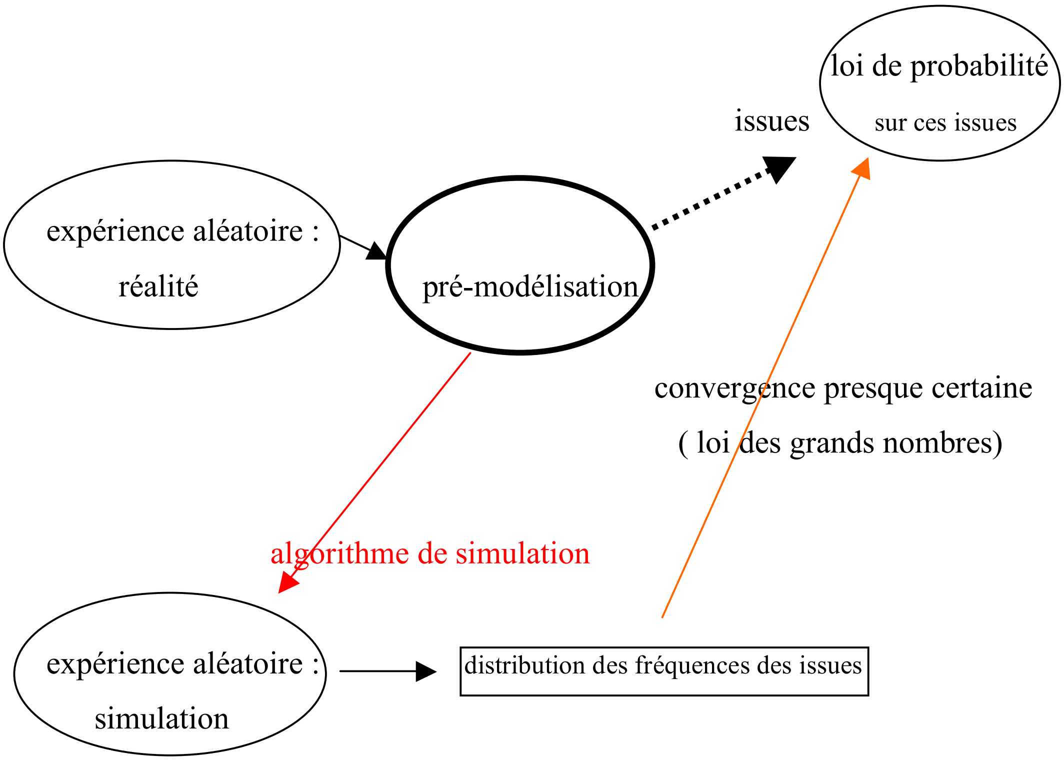
on ne possède pas de données expérimentales pour l'expérience initiale :



Le test du χ^2 peut permettre de valider la simulation à un risque d'erreur donné. En théorie, il permet de dire qu'à un risque d'erreur donné, le test n'invalide pas la simulation.

Schéma 6

on n'a pas de modèle pour l'expérience initiale :

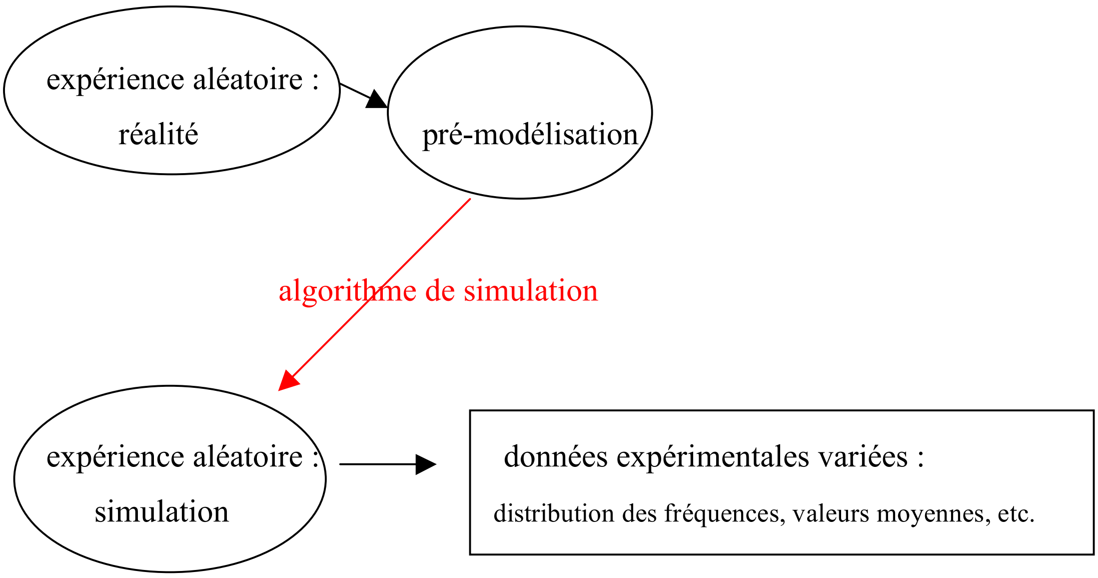


La pré-modélisation amène à déterminer les *hypothèses de modélisation* pour les expériences aléatoires élémentaires qui composent l'expérience étudiée, hypothèses qui seront communes à la modélisation probabiliste et à la simulation ; les issues retenues dans le modèle probabiliste à construire seront les issues de la simulation, que l'on relèvera en répétant cette simulation un grand nombre de fois. Une étude statistique des résultats ainsi observés permettra de calculer la distribution des fréquences de chacune des issues, distribution des fréquences qui tend en probabilité vers la loi de probabilité de l'expérience aléatoire initiale.
La simulation est alors un procédé pour construire une loi de probabilité raisonnable de l'expérience initiale. Cela présente particulièrement un intérêt si l'expérience initiale n'est pas (facilement) reproductible, et donc ne peut fournir de données empiriques exploitables.

Revenons au cas où l'on ne possède ni données expérimentales, ni loi de probabilité, auxquelles confronter la simulation. On se trouve dans la situation présentée par le schéma 7.

Schéma 7

La simulation peut permettre de prendre des décisions :



3. Retour sur les deux étapes : pré-modélisation et création d'un algorithme

La pré-modélisation est une étape qui comporte une phase de réduction (concernant essentiellement les issues des expériences aléatoires élémentaires composant l'expérience globale) et une phase de choix d'*hypothèses de modélisation* sur chacune de ces expériences élémentaires. Il n'y a pas unicité de ces choix. C'est ce qui amène la multiplicité des modèles possibles.

Ces choix se font « au coup par coup » en fonction du contexte, probablement avec certaines heuristiques ; il n'est pas dans notre propos d'étudier les procédures en jeu, dans la mesure où dans l'enseignement secondaire elles paraissent relativement transparentes.

La « mise en algorithme » est une tâche qui suppose qu'on ait pu décrire l'expérience aléatoire en termes d'actions qui s'enchaînent, ce que nous avons appelé *le cahier des charges* associé à l'expérience.

Prenons comme exemple l'expérience aléatoire suivante :

Une marque de chocolat glisse dans les tablettes des images. Il y a six images différentes. Combien faut-il acheter de tablettes de chocolat pour espérer avoir la collection complète ?

On peut décider, pour résoudre ce problème, d'essayer de faire une simulation.

Cela suppose d'abord de formuler plus précisément la description du problème pour faire ressortir qu'il met en jeu une expérience aléatoire :

Une marque de chocolat glisse une image dans chaque tablette de chocolat. Il y a six images différentes que l'on peut coller dans un « collector ». Lors de l'achat d'une tablette de chocolat, on ne peut pas savoir quelle est l'image qui figure dans la tablette. L'image que l'on trouve dans la tablette est donc aléatoire. On veut simuler l'expérience aléatoire qui consiste à acheter des tablettes de chocolat tant que la collection des six images n'est pas complète.

Ce qui est dans l'encadré est le cahier des charges à partir duquel pourront être décrites les différentes actions à simuler.

Commençons par l'étape de pré-modélisation :

• On peut numéroter les types d'images : 1, 2, …, 6. Au fur et à mesure des achats de tablette de chocolat, on constitue la liste des numéros déjà collectés. La collection est complète lorsqu'on a obtenu la liste {1 ; 2 ; … ; 6}. Le numéros d'image que l'on découvre lorsqu'on ouvre une tablette que l'on vient d'acheter est le résultat de l'une des expériences aléatoires élémentaires qui composent l'expérience globale : on achète des tablettes jusqu'à avoir la collection complète.

Les issues de chacune des expériences aléatoires élémentaires sont les numéros 1, 2, …, 6.

• Quelles *hypothèses de modélisation* faire ?

- Si l'on imagine que la chocolaterie ne cesse de produire des tablettes de chocolat en mettant en circulation, à tout moment, autant d'images de chaque sorte, l'ensemble des tablettes produites étant « brassé » avant répartition chez les distributeurs, les numéros suivront une loi équirépartie, et lors du deuxième achat, la probabilité de découvrir l'image numéro 3 sera la même que lors du premier achat : 1/6 ; les achats sont deux à deux indépendants : il s'agit ici d'une *indépendance a priori* (Maury (1985))
- Si l'on imagine que la chocolaterie a produit 30 tablettes de chocolat avec 5 images de chaque type, la probabilité de découvrir l'image numéro 3 lors du deuxième achat dépendra du numéro découvert lors du premier achat : 5/29 ou 4/29.
- Plus généralement, si l'on imagine que la chocolaterie a produit $6n$ tablettes de chocolat (n tablettes pour chacun des 6 types d'image), la probabilité de découvrir l'image numéro 3 lors du deuxième achat dépendra du numéro découvert lors du premier achat : $(n-1)/(6n-1)$ ou $n/(6n-1)$. Si n est très grand, ces deux nombres seront très voisins de 1/6, et on aura des hypothèses qui amèneront un modèle voisin de celui qu'on peut imaginer pour le premier

choix.

Voilà pour l'étape de pré-modélisation.

Essayons maintenant de décrire les actions qui constituent l'expérience aléatoire que constitue le fait d'acheter des tablettes de chocolat jusqu'à obtenir la collection complète :

- On se procure un « collector » pour ranger les images. (1). Le collector au départ ne contient pas d'image.
- On achète une première tablette et on récupère l'image qu'elle contient ; on colle cette image à sa place dans le collector. (2)
- On achète une seconde tablette et on récupère l'image qu'elle contient. On consulte le collector : si l'image ne s'y trouve pas encore, on la colle, à sa place, dans le collector. (3)
- On continue ainsi d'acheter des tablettes comme on vient de le décrire, tant que le collector n'est pas plein. (4)

Il est important non seulement de bien décrire les actions, mais aussi de bien décrire leurs enchaînements, en particulier ce qui se passe au départ (on n'a rien), et quand ça s'arrête (quand la collection est complète).

Cette description des actions et de leur enchaînement est la conséquence du *cahier des charges* ; elle va se traduire par un algorithme :

Initialiser un tableau T de longueur 6 : **Pour i allant de 1 à 6, T(i) ← 0**	(1) : le tableau T est le collector ; chaque case de T est au départ à 0 : aucune image n'est à sa place, le collector est vide
Tant que la somme des éléments de T ne vaut pas 6	(4) : Tant que la collection n'est pas complète
Choisir un entier aléatoire X selon le choix d'hypothèses défini dans la pré-modélisation **Si T(X) = 0 alors T(X) ←1**	(2) et (3) : On achète une nouvelle tablette et on découvre l'image ; Si l'image n'est pas déjà dans le collector, on la range à sa place dans le collector
Fin de Tant que	

Remarquons que la mise en œuvre de cet algorithme est une simulation de l'expérience aléatoire, mais qu'elle ne produit pas de données expérimentales.
On peut compléter le cahier des charges :

Une marque de chocolat glisse une image dans chaque tablette de chocolat. Il y a six images différentes. Lors de l'achat d'une tablette de chocolat, on ne peut pas savoir quelle est l'image qui figure dans la tablette. L'image que l'on trouve dans la tablette est donc aléatoire. On veut simuler l'expérience aléatoire qui consiste à acheter des tablettes de chocolat tant que la collection des six images n'est pas complète ***et répéter 100 fois cette simulation afin de calculer combien, en moyenne, il faut acheter de tablettes de chocolat pour que la collection d'images soit complète****.*

il faut alors ajouter à l'algorithme tel qu'il vient d'être décrit un compteur de tablettes achetées, et insérer l'algorithme de simulation à proprement parler de l'expérience aléatoire, dans une itération qui permettra de calculer la valeur moyenne demandée. Ces rajouts sont des activités algorithmiques de base pour un informaticien, mais demandent un apprentissage pour un élève de lycée.

Les algorithmes à mettre en œuvre pour simuler au lycée des expériences aléatoires, sont simples. Cependant leurs constructions supposent des connaissances ; c'est pourquoi nous allons faire dans le chapitre suivant un détour sur l'algorithmique-programmation et son apprentissage en lycée.

CHAPITRE 10

L'enseignement de l'algorithmique-programmation

Ce chapitre ne fait qu'évoquer quelques questions liées à l'enseignement de l'algorithmique-programmation.
Certaines de ces questions ont pu apparaître en filigrane dans les chapitres précédents (en particulier au chapitre 6), d'autres seront plus nettement explicitées dans des chapitres ultérieurs, en particulier au chapitre 11, et au chapitre 14 paragraphe 3.
Concernant l'aspect « informatique » des simulations, nous avons pu nous appuyer, au fil de cette thèse, sur des documents très variés : aussi bien des ouvrages de référence concernant la programmation que des articles à caractère plus ponctuel, concernant l'enseignement de l'informatique au lycée dans les années 1980-1990, ou concernant l'utilisation de l'ordinateur en classe de mathématiques.
Nous renvoyons tout particulièrement, pour plus de développements sur l'enseignement de l'informatique et plus précisément l'enseignement de l'algorithmique-programmation, à plusieurs ouvrages ou articles : Arsac (1977, 1987, 1980, 1988), Baron (1987, 2001), Colloque IFIP-OCDE (1970), Dijkstra (1968, 1972, 1976), Naur (1969), Knuth (1968), Gries (1981), Pair (1989), Rogalski (1986 à 1989).
Nous aurons l'occasion de citer la plupart de ces auteurs dans ce chapitre, et au delà.

L'informatique a été enseignée dans l'enseignement secondaire <u>général</u> (c'est à dire en lycée *classiques* ou *polyvalents*) pendant une décennie et demi, dans le cadre d'un enseignement optionnel d'informatique à visée non professionnelle. On peut trouver une chronologie des « événements » qui ont marqué cette option de sa naissance à sa disparition élaborée par l'association E.P.I. : document numérique [4] de la bibliographie.
L'informatique a également été enseignée dans les lycées techniques, avec une perspective de formation professionnelle.

Il y avait dans les programmes de l'option informatique de l'enseignement général une part importante d'algorithmique-programmation ; le savoir-faire visé était : écrire un algorithme permettant la résolution d'un problème, le traduire dans un langage de programmation et faire exécuter ce programme par un ordinateur.

Une telle tâche relève d'un type de tâche sur lequel il y eut des recherches didactiques pendant cette période, que nous appellerons dans la suite période de *l'option informatique*.
Ces recherches ont eu une double particularité :
• elles ont été menées en même temps que se constituait la discipline informatique dans l'enseignement secondaire général ;
• elles associaient directement les enseignants de lycée engagés dans cet enseignement, enseignants qui n'étaient pas a priori « didacticiens », mais se trouvaient confrontés à divers problèmes didactiques pour lesquels personne n'avait encore d'expérience : l'enseignement universitaire était lui-même très jeune, et personne ne pouvait s'appuyer sur sa propre expérience d'élève.

Les compétences en algorithmique-programmation étaient acquises à l'origine sur le terrain de l'entreprise, avant de devenir un objet d'enseignement universitaire, puis scolaire.

Pour des raisons que nous évoquerons dans ce chapitre, l'enseignement *es qualité* de l'algorithmique-programmation dans l'enseignement général des lycées a été abandonné dans les années 1990.
Figurent pourtant dans les préambules des programmes de mathématiques depuis cette période, y compris les « programmes 2000 » une description des compétences que les élèves scientifiques doivent posséder, et mettre en œuvre [58], sans qu'il soit jamais précisé comment ils les acquièrent. Cette partie de la prescription est restée en grande partie lettre morte, entre autres raisons parce que la majorité des enseignants de mathématiques ne se sentait pas qualifiée pour assumer cet enseignement.

Il apparaît que, les simulations informatiques d'expériences aléatoires dans l'enseignement secondaire relevant aussi de l'algorithmique-programmation, il est nécessaire de faire ici un détour sur des questions didactiques spécifiques à l'algorithmique-programmation.

---

[58] « Certaines notions informatiques élémentaires (boucle, test, récursivité, tri, cheminement dans des graphes, opérations sur les types logiques) font partie du champ des mathématiques et pourraient être objets d'enseignement dans cette discipline. Compte tenu de l'horaire imparti et des débats en cours, il n'est proposé ici aucun chapitre d'informatique. Néanmoins, l'élève devra mettre en œuvre, notamment su sa calculatrice, les notions de boucles et de test. » Mathématiques, Série scientifique, Enseignement obligatoire, B.O. n° 7 hors série, 31 août 2000, p. 169.

1. De l'algorithmique-programmation aux T.I.C.E., et réciproquement.

La nécessité d'une « alphabétisation » des élèves en informatique s'est imposée dans les années 1980, du fait de la possibilité (toute relative d'ailleurs) pour les établissements scolaires d'acquérir des ordinateurs (souvent : un ordinateur) ; les initiatives étaient renforcées par une volonté politique de soutenir la « filière française » (plan informatique pour tous, et micro-ordinateurs Thomson).

Assez vite cependant, on a vu s'opposer deux tendances :

- d'une part l'idée qu'une utilisation en citoyen éclairé de l'outil informatique passe par une compréhension du fonctionnement d'un ordinateur, autant de sa partie matérielle que de sa partie logicielle ; cette conception s'est traduite à la rentrée scolaire 1980 par la mise en place d'une « option informatique » en classe de seconde à titre expérimentale dans douze lycées, enseignement élargi aux classes de premières et terminales dans les années suivantes, mais également élargi à d'autres lycées selon une croissance importante: 550 lycées en 1986 (source : E.P.I.) ;
- d'autre part l'idée que l'initiation à l'outil informatique n'avait pas à se faire dans un enseignement spécifique, qu'elle serait portée d'elle-même par les diverses disciplines qui ne pourraient pas se passer d'ordinateurs comme outils accompagnant l'enseignement de la discipline.
    - Cet accompagnement pouvait être sous la forme d'E.A.O. (enseignement assisté par ordinateur) : on a vu ainsi un didacticiel d'entraînement au résumé de texte pour l'enseignement des lettres, ou des didacticiels d'entraînement au calcul (mental aussi bien qu'algébrique). Certains didacticiels ont pu, dans les années 1990, être conçus en liaison avec des recherches didactiques disciplinaires : on a parlé alors d'E.I.A.O. (enseignement intelligemment assisté par ordinateur).
    - Cet accompagnement pouvait aussi, particulièrement dans les disciplines scientifiques, être l'utilisation de l'ordinateur comme outil de calcul (tableurs, logiciels de mesures physiques), comme outil de construction géométrique (les logiciels Cabri-géomètre ou Géoplan/Géospace), ou comme outil intégré à la formation professionnelle (tableurs et traitement de texte dans les formations du tertiaire, logiciels d'aide à la conception dans des formations techniques).

La première tendance s'est imposée dans les années 1980 dans l'enseignement général des lycées, probablement du fait de la modestie des équipements (les lycées généraux, ne bénéficiant pas de la taxe professionnelle, ne pouvaient compter que sur des dotations ministérielles, puis régionales), mais également du fait du manque de logiciels (*progiciels* ou *didacticiels*) adaptés en tant qu'outils accompagnant les enseignements disciplinaires.

L'enseignement de l'informatique en tant que discipline à part entière s'est donc développé, très vite dans l'enseignement général, peut-être d'autant plus vite que les débouchés en terme d'emplois semblaient immenses : les élèves et leurs familles pensaient prendre une assurance sur l'avenir en choisissant cette option (qui fut dès l'année 1884 sanctionnée par une épreuve au baccalauréat au même titre que les autres enseignements optionnels).

Cependant, les métiers de l'informatique ont évolué très vite, les « programmeurs » disparaissant au bénéfice des « analystes programmeurs » ; les langages de programmation ont évolué aussi, exigeant très vite des « recyclages ».
Cela a pu faire douter de l'existence d'un noyau de connaissances scientifiques stable attaché à l'informatique.
Il y eut également une crise dans le recrutement des enseignants scientifiques, et il est un fait que l'enseignement de l'option informatique en pleine expansion était assuré essentiellement par des professeurs de mathématiques et de sciences physiques, qui faisaient alors défaut dans leurs disciplines d'origine.
Dans certains lycées « polyvalents » (c'est à dire possédant à la fois des séries générales et des séries technologiques du secteur tertiaire), cette option semblait concurrencer d'autres enseignements, professionnels, qui utilisaient des ordinateurs.

Les programmes de cette option informatique comprenaient tout à la fois une partie d'étude du matériel, une partie d'algorithmique-programmation, et une partie « informatique et société » (la loi dite « *informatique et libertés* » a été votée en 1978).
La formation des enseignants était en moyenne d'un an (soit en décharge complète : ce furent les « stages lourds » d'informatique, qui prenaient également en charge l'utilisation de l'outil informatique dans les disciplines, soit dans le cadre du plan de formation académique), formation qui s'ajoutait à la formation dans la discipline d'origine (au minimum la licence pour les professeurs certifiés) ; certains enseignants possédaient une formation universitaire

en informatique de niveau licence ou maîtrise.
Les tentatives de mettre en place une Inspection Générale d'informatique ont échoué.

Un tel enseignement n'avait, bien sûr, aucune tradition dans l'enseignement secondaire, pas plus à dire vrai qu'à l'université : les départements d'informatique sont restés longtemps des appendices d'UER scientifiques, les enseignants venant d'horizons divers.

Il semble que certains lycées aient laissé dériver le contenu de la prescription vers la seule algorithmique-programmation. A la fin des années 1980 un rapport annonçait « la fin programmée » de cette option :

> *Nous ne pensons pas que l'informatique doive être enseignée comme discipline (théorique) en tant que telle à ces niveaux de formation. L'informatique enseignée à ce niveau contient des risques de formalisme encore plus graves que les mathématiques. L'argument des élèves mauvais en mathématiques qui se rattraperaient par l'informatique est peu fondé. Par contre l'introduction d'outils informatiques peut « débloquer » des élèves en difficulté et en motiver d'autres.* (Rapport de la mission Dacunha-Castelle,1989)

La seconde tendance, celle d'une acquisition de connaissances informatiques par l'utilisation d'outils informatiques, a été imposée, plus qu'elle ne s'est imposée, par la suppression de l'option informatique.
La volonté politique d'une utilisation d'outils informatiques dans les disciplines s'est parfois exprimée de façon maladroite autant que volontariste (quantification du nombre de séances par semaine en salle informatique par certaines Inspections Générales disciplinaires par exemple, ou introduction d'épreuves pratiques au baccalauréat en sciences physiques, en SVT, et en mathématiques, cette dernière épreuve n'ayant encore existé qu'à titre expérimental).

Cette volonté s'est heurtée à une certaine inertie des enseignants.
Certains ont voulu y voir une volonté de continuer à enseigner « comme avant », et ont dénoncé un corps enseignant passéiste.
Il y a peut-être eu, en réalité, une insuffisante prise en compte de la nécessité d'une formation conséquente et cohérente des enseignants.

Quelques stages ont certes été dispensés.

Mais il n'a jamais été véritablement pris en compte que l'utilisation de l'outil informatique en classe exige deux choses étroitement liées :

- que l'enseignant ressente l'utilisation de l'ordinateur comme une nécessité dans sa progression pédagogique,
- qu'il puisse préparer cette progression dans de bonnes conditions.

Alors même que l' « Entreprise » dotait ses cadres d'ordinateurs portables, l'enseignant disposait, dans le meilleur des cas, d'un ordinateur dans une hypothétique salle de travail dans son lycée, à partager avec de nombreux collègues.

Les précurseurs de l'utilisation des TICE dans leur classe ont probablement été des enseignants qui se sont équipés sur leurs deniers d'un ordinateur personnel, et qui ont pu mettre à profit chez eux les formations glanées au fil des stages.

On peut penser que le taux d'équipement en ordinateur familial était le même dans les foyers d'enseignants que dans les foyers de revenus équivalents.

Les coûts des matériels informatiques baissant, les familles se sont équipés d' « ordinateurs domestiques ».

Cependant, les avancées de la technologie et la baisse des coûts du matériel informatique ne se sont pas immédiatement traduites, pour les enseignants, en termes d'outils incontournables, que ce soit pour « faire cours », ou pour la préparation de ces cours (traitement de texte, carnet de notes informatique).

Nous avons donc vu une politique volontariste peu suivie d'effets : ainsi la création du « B2I » des collèges, est restée en général lettre morte, jusqu'à ce que ce « B2I » devienne un élément de la note pour obtenir le brevet des collèges.

Le « B2I » a été un temps porté par l'enseignement de la technologie en collège, jusqu'aux récents changements des programmes de cette discipline.

L'organisation actuelle du « B2I » (qui comporte des savoir-faire précis) de façon « transdisciplinaire » semble poser des problèmes : qui enseigne quoi, et qui évalue quoi ?

Les T.I.C.E. (technologies de l'information et de la communication pour l'éducation) entrent petit à petit dans les classes de lycée, dans des disciplines aussi variées que les disciplines scientifiques, l'histoire-géographie ou le grec ancien.

Et qu'en est-il de l'algorithmique-programmation ?
Comme nous l'avons déjà souligné, les programmes de mathématiques comportent depuis les années 1990 des compétences en matière d'algorithmique-programmation allant jusqu'à la capacité à faire une boucle avec un test d'arrêt, compétences jamais évaluées au baccalauréat et le plus souvent oubliées dans le « savoir enseigné ».

Le rapport Kahane [59] évoque les besoins de la société vis à vis d'une formation en probabilités et statistiques, et soulève à nouveau la question de l'enseignement de l'informatique en tant que discipline :

> *« Les programmes de mathématiques des années 90 stipulent que l'élève devra maîtriser l'usage d'une calculatrice scientifique (seconde et première) et d'une calculatrice programmable (terminale). Si cette requête est affichée dans l'introduction du document, il n'en est plus fait mention dans l'évocation des contenus et des activités. Comment le législateur imaginait-il cet apprentissage ? Il est peut-être temps de faire entrer ce souhait dans les faits en s'en donnant les moyens ».* (Rapport de la commission Kahane,2001)

Les auteurs des programmes scientifiques des années 2000 n'ont pu aboutir à un accord sur ce point.
En revanche, les programmes de mathématiques des années 2010 intègrent des éléments d'algorithmique-programmation conséquents en classe de seconde (et pas seulement affichés en introduction des programmes, comme le relevait la commission Kahane).
Et surtout, une spécialité informatique (plus précisément : « informatique et sciences du numérique » , I.S.N.) est annoncée pour la rentrée 2012 en terminale scientifique, au même titre que les spécialités mathématiques, physiques, et sciences de la vie et de la terre.

La prescription de simulations informatiques d'expériences aléatoires dans les programmes de mathématiques de lycée des années 2000 puis 2010 doit donc être replacée dans ce contexte :

---

[59] Ce rapport est disponible sur le site de la SMF (Société de Mathématique de France) http://smf.emath.fr/Enseignement/CommissionKahane/

• à l'origine, utilisation des T.I.C.E. par l'enseignant pour introduire ou soutenir l'enseignement de notions du programme de probabilité, l'enseignant étant l'acteur principal des simulations informatiques ; le programme de seconde évoque l'utilisation par l'élève de la touche « rand » de la calculatrice, tandis que le programme de terminale S évoque des simulations « extérieures » générant des résultats statistiques communiqués aux élèves pour traiter de l'adéquation de la loi équirépartie aux données expérimentales ;
• puis, avec l'introduction d'une épreuve pratique (expérimentale) de mathématiques au baccalauréat scientifique, création **<u>par les élèves</u>** de simulations sur tableur, le tableur étant supposé exiger peu de connaissances en algorithmique programmation.
• Et enfin création par les élèves de simulations d'expériences aléatoires dans un langage impératif : les élèves sont initiés dès la seconde à l'algorithmique.-programmation

Ni les documents d'accompagnement des programmes des années 2000, ni les documents d'accompagnement du programme de seconde de la rentrée 2009, ni les programmes (2000 ou 2010) n'évoquent d'éventuels problèmes didactiques concernant ce type d'activité.

Nous limiterons notre propos à certains aspects qui « risquent » d'apparaître lorsqu'on met des élèves en situation de **<u>créer</u>** des simulations informatiques d'expériences aléatoires.
Nous nous centrerons sur la difficulté à construire un algorithme, et particulièrement un algorithme comportant une itération, en nous appuyant sur deux contributions au « Colloque francophone sur la didactique de l'informatique » qui s'est tenu à Paris en 1988 : contribution de Jacques Arsac, professeur d'informatique à l'université, et de Janine Rogalski, chercheur en didactique.

2. « La didactique de l'informatique : un problème ouvert ? »

C'est le titre de la contribution d'Arsac (1988) à ce colloque, qui énonçait :

> *« Dès le début des années 60, les professeurs qui enseignaient les ordinateurs, leur structure et leur architecture, leur programmation, leurs applications dans les domaines numériques et non numériques prirent conscience qu'une science nouvelle était en gestation (computer science aux USA, datalogie en Scandinavie, informatique en France) et que cela posait de sérieux problèmes d'enseignement. Quels étaient les contours de cette science ? Que fallait-il enseigner ? A qui ? Avec quels prérequis ? »*

Arsac décrit brièvement la constitution de la science informatique : à la fin des années 1960, « *la programmation était un empirisme* », ou, selon Knuth (1968), un art.

L'empirisme ici consiste pour le programmeur à construire son programme de façon essentiellement intuitive et personnelle, en comptant sur l'exécution d'un certain nombre de jeux de données où le programme fonctionnera correctement (c'est à dire fournira les résultats attendus pour ces données), pour **se** convaincre que le programme est juste, c'est à dire fait bien ce pour quoi il a été commandé.

Les défauts des logiciels construits par empirisme sont rapidement devenus socialement trop coûteux : ces logiciels étaient non fiables et non modifiables entre autres défauts.

Des chercheurs avaient posé la question des preuves de programme. D'autres (ou les mêmes) essayent de dégager des méthodes scientifiques de programmation : *programmation descendante* par Naur (1969) et *programmation structurée* par Dijkstra (1972).

Arsac poursuit :

> *A la suite de ces travaux de recherche, la programmation changea de nature. Elle approcha du statut de discipline scientifique (A discipline of programming, Dijkstra [DIJ76]), puis fut reconnue comme science (The science of programming, Gries [GRI81]). Ces méthodes scientifiques changèrent la façon de programmer, et l'on vit apparaître la "méthodologie de la programmation" comme champ de recherche, avec des méthodes de programmation comme résultat concret. Les premières naquirent dans le cadre relativement restreint de l'informatique de gestion [WAR75], [JAC75]. Des méthodes de portée plus générale, et partant moins précises dans leurs détails, furent proposées pour aider à la création de nouveaux algorithmes ([DIJ76], [ARS77], [ARS80], [GRI81], [DUC84]). Tout ceci ne pouvait manquer d'avoir un impact sur l'enseignement.*

Dans la même période, des domaines professionnels de plus en plus nombreux se mettent à utiliser l'ordinateur. Se pose alors la question : les professionnels, qu'ils soient ingénieurs, mathématiciens, économistes ou géographes doivent-ils écrire eux-mêmes les logiciels dont ils ont besoin, ou doivent-ils faire confiance à des bibliothèques de programmes que d'autres ont écrits ? Quels moyens ont-ils pour contrôler les résultats ?

Faut-il, demande Arsac en 1989, former les élèves ?

> *« L'élève doit-il savoir quelque chose de l'ordinateur et de la façon dont on lui fait réaliser une tâche, ou peut-on considérer matériel et logiciel comme une boîte noire ? Quelle compétence est demandée au professeur ? Doit-il être un expert en traitement de texte, tableur ou gérant de bases de données? Doit-il savoir de l'informatique? Doit-il être compétent en programmation ? »*

On pouvait encore en 1989 se demander si les connaissances informatiques seraient stables dans le temps ; l'histoire de l'informatique avait été tellement vite :

- augmentation de la capacité de la mémoire utilisable dans l'unité centrale,
- augmentation de la capacité de la mémoire de masse,
- multiplicité (et en même temps standardisation) des supports de la mémoire de masse,
- augmentation de la vitesse de calcul,
- et surtout développement de langages de programmation de plus en plus ouverts, vers une programmation de plus en plus près de la pensée du programmeur (et partant plus éloignée du langage machine), c'est à dire de plus en plus proche de l'algorithme tel que le programmeur le conçoit pour résoudre un problème selon un cahier des charges donné, avec divers styles de programmation, selon le type de problème à résoudre, ou l'outil de programmation à disposition.

Il apparaissait que, quelle que soit l'évolution de l'informatique, une formation de base et de masse était nécessaire, pour que les usagers de l'ordinateur soient des usagers éclairés.
Citons encore Arsac :

> *« Tout ceci pose de manière grave le problème du contenu des enseignements d'informatique pour tout public autre que les futurs professionnels de cette discipline (en admettant que pour eux la question soit résolue, ce qui paraît tout de même une hypothèse raisonnable). Faut-il apprendre aux gens le maniement d'outils dont on connaît la rapide obsolescence ? Peut-on estimer que des fonctions essentielles ont été mises en évidence qui perdureront dans les futurs produits, et qu'à travers les réalisations actuelles, ce sont elles que l'on enseigne ? Est-ce bien cela qui est actuellement pratiqué ? Faut-il au contraire renoncer à l'idée de boîte noire, jusqu'où faut-il aller ? Quelle quantité de programmation est nécessaire pour comprendre ce que peuvent, et peut-être ne peuvent pas les ordinateurs ? »*

Voilà clairement posée la question de la transposition didactique des connaissances d'une discipline scientifique toute neuve, et qui évolue à une vitesse inédite pour les disciplines enseignées jusque là.

Concernant la programmation, Arsac écrit en 1988 :

> *« Il semble à peu près compris que programmer n'est pas résoudre un problème, c'est le faire résoudre par une machine. Les élèves ou les étudiants savent résoudre la plupart des problèmes qu'ils auront à programmer, qu'il s'agisse de trouver quel jour de la semaine tombe un jour donné, ou d'écrire en lettres un montant donné en chiffres, ou de trier une suite de nombres. Or pour faire faire un travail par un exécutant qui n'a pas été construit de façon spécifique pour ce travail là (les ordinateurs sont des machines universelles à traiter l'information), il faut fournir une méthode à cet exécutant. On est donc confronté au problème suivant : formuler la méthode de résolution d'un problème que nous savons résoudre.*
>
> *Certains pensent que la difficulté est dans la formalisation de la méthode : nous savons faire, mais nous savons mal dire comment nous faisons. On essaie alors d'expliciter notre propre façon de faire. On manifeste l'enchaînement de nos actions au moyen d'un organigramme, puis on code le résultat dans un langage de programmation : l'analyse informatique part d'une analyse du vécu. Il y a des cas où c'est efficace (...)*
>
> *De nombreux ouvrages sont fondés sur cette démarche pédagogique.*

Arsac explique qu'en réalité, on ne programme pas la résolution d'un problème donné à partir d'une explicitation des procédures que l'acteur humain met en œuvre pour résoudre le problème « à la main », procédures qu'il suffirait de transcrire dans un langage exécutable par la machine. Il illustre cette affirmation avec les programmes de tris : la résolution d'un problème de tri « à la main » se fait par un assemblage plus ou moins désordonné de procédures de tris locaux, que l'acteur adapte au fur et à mesure de l'avancement de la tâche. Arsac conclut sur ce point :

> *« Pourquoi opéreraient-ils de façon systématique ? Il suffit qu'ils réussissent .. .*
>
> *En réalité, on a une méthode d'enseignement qui part du programme à écrire, puis le paraphrase en français courant en disant que c'est comme cela que nous faisons. C'est une duperie inacceptable.*
>
> *(...)*
>
> *L'enseignement de la programmation se heurte à cette difficulté fondamentale : comment fabriquer une méthode pour résoudre un problème? On appelle cela*

"l'algorithmique". Le terme me paraît abusif : l'algorithmique est l'étude des algorithmes, de ce qu'ils permettent de calculer, de la façon dont on peut les composer, les prouver, mesurer leur complexité, ... Ce n'est pas cela qui est en cause, c'est bien moins ambitieux : seulement trouver, face à un problème donné, une méthode qui en viendra à bout. Il faut développer la créativité des élèves ou des étudiants.

Il n'y a pas de réponse universelle. Il n'y a pas un algorithme pour la fabrication d'algorithmes. Il y a des façons de faire. Il y a des méthodes de raisonnement plus adaptées que d'autres : la programmation impérative aussi bien que la programmation récursive se fondent sur le raisonnement par récurrence. Il est important de le savoir, parce que cela fournit un cadre intellectuel. Mais la difficulté demeure: comment trouve-t-on une hypothèse de récurrence qui deviendra le coeur de la procédure récursive, ou l'invariant de boucle pour un programme impératif.

3. Enseignement de méthodes de programmation dans l'initiation à l'informatique

Jacques Arsac était professeur d'informatique à l'université Paris VII, après avoir été chercheur en physique. Il a su, comme on vient de le voir, très vite réaliser que l'enseignement de l'informatique, dans sa partie programmation, posait des problèmes didactiques spécifiques. On pourrait dire qu'il a été l'un des acteurs en France qui a permis que la programmation passe d'une pratique essentiellement empirique, à une activité scientifique. Son souci premier était que les élèves de lycée, s'ils devaient être initiés à la programmation, le soient d'emblée à une « bonne » programmation : « mon programme est juste parce que je l'ai construit par des méthodes qui m'assurent qu'il est juste ».

L'une de ces méthodes concerne la construction d'une itération à partir d'une « situation », c'est à dire d'une hypothèse supposée vraie en un point donné de la boucle, à chaque passage dans la boucle. Cette hypothèse s'énonce dans les termes : « supposons que j'ai déjà fait une partie du travail », hypothèse qu'on décline au cas par cas, en l'adaptant au problème concret à résoudre. Après avoir regardé si le travail est terminé (« est-ce fini ? », question qui amène à définir un booléen qui sera le critère d'arrêt de la boucle), il s'agit alors de déterminer les

actions à effectuer (i.e. : les instructions à introduire dans le programme), pour « avancer le travail ». Il reste enfin à déterminer les valeurs initiales des variables pour que l'hypothèse soit satisfaite au premier passage dans la boucle (étape d'initialisations des variables).
Evidemment cette méthode doit être rapprochée du raisonnement par récurrence, qui est généralement enseigné dans les seules classes scientifiques, au niveau de la terminale.
La *situation* est ce qu'on appelle un i*nvariant de boucle*, notion qui peut être un élément de preuve du programme.

Rogalski (1989) a analysé cette méthode dans un cadre de recherche didactique en informatique.

> *« L'enseignement de l'informatique à des élèves (ou des étudiants), qui ne sont pas a priori de futurs professionnels de l'informatique, pose des problèmes didactiques aussi épineux que ceux rencontrés par les enseignants des différentes disciplines "traditionnelles". L'informatique y apparaît à la fois comme un objet, relevant d'un domaine scientifique avec ses concepts propres, et comme un outil pour contribuer à résoudre les problèmes d'autres domaines.*
> *Parmi ses spécificités - du point de vue du contenu et du point de vue des élèves [60]- celle d'être une discipline de service, liée à de multiples pratiques professionnelles, a (indirectement ?) conduit au problème de l'enseignement de méthodes de programmation. La réflexion sur l'efficacité du travail des programmeurs professionnels a en effet convergé, vers la fin des années soixante, avec des approches de nature formelle et les travaux sur la programmation structurée (Dahl, Dijktra, Moare, 1972). »*

La note (cf. au bas de cette page) sur la diversité des transpositions didactiques, provenant à la fois des connaissances d'une discipline scientifique nouvelle, et de pratiques professionnelles, mérite d'être soulignée : ont coexisté dans l'enseignement secondaire un enseignement professionnel enseigné dans les lycées techniques, débouchant sur un baccalauréat technique

[60] La question des spécificités de l'informatique est liée étroitement au problème des diverses transpositions didactiques qui peuvent être utilisées pour passer d'une connaissance du domaine scientifique ou des pratiques professionnelles à un objet d'enseignement. Les deux types de transposition sont également en relation avec les représentations qu'ont les enseignants sur l'informatique et ce qu'ils souhaitent que se représentent les élèves. Le point de vue des spécificités de l'informatique est développé dans (Rogalski, 1987a).

spécifique (le bac H), et un enseignement optionnel de l'informatique dans les lycées d'enseignement général, qui se voulait ouvert à toutes les séries (aussi bien C et D que A2 ou A1, ou que B et G). Rogalski poursuit, à propos de l'enseignement de méthodes de programmation :

> *« Cela a conduit à concevoir la structure d'un programme de façon hiérarchique et modulaire, à élaborer des méthodes de programmation dans le domaine professionnel et à engager une réflexion sur les conséquences à en tirer dans le domaine de l'enseignement de la programmation, indépendamment de la perspective professionnelle.*
>
> *L'hypothèse didactique était "qu'une programmation méthodique dès les premiers apprentissages évite de prendre des habitudes de programmation empirique et désordonnée, tout en facilitant la résolution des problèmes de programmation" (Rogalski, Samurçay, Hoc).*
>
> *Plusieurs années d'expérience d'enseignement de l'informatique dans l'option des lycées conduisent à penser que les difficultés d'enseignement de méthodes ont été sous-estimées. Avant de pousser à des révisions déchirantes (C. Pair,1988), il n'est pas inutile d'analyser plus précisément le ou les objets d'enseignement relatifs aux méthodes de programmation pour essayer d'identifier les propriétés de situations d'enseignement propres à faire acquérir par les élèves cette approche méthodique reconnue comme nécessaire. »*

Rogalski souligne d'autres difficultés spécifiques à la discipline informatique :

> *« Un premier obstacle apparaît dans l'acquisition de l'itération, et de méthodes appropriées à l'écriture de programmes itératifs ou récursifs. On peut exprimer cet obstacle sous deux formes différentes. D'une part la maîtrise de la dialectique entre exécution d'actions (ou "calculs") avec son caractère dynamique, et succession de situations (ou "relations") avec un caractère statique de la situation semble au coeur de l'acquisition des constructions itératives ou récursives.*
>
> *D'autre part, la variable informatique doit être conçue comme une fonction de l'exécution ; une telle dualité variable/fonction n'est pas spécifique à l'informatique mais elle n'est jamais apparue dans le "passé cognitif" des élèves (qui maîtrisent en général fort mal à l'issue de l'enseignement secondaire les*

relations variable/fonction). La notion d'invariant qui est au coeur de la présentation de l'itération par J. Arsac peut difficilement recevoir une signification sans la maîtrise de cette double perspective. »

Nous avons pensé utile ce bref aperçu sur quelques problèmes didactiques soulevés lors de l'enseignement de l'informatique en lycée comme une discipline en tant que telle, enseignement dispensé dans les années 1980/1990.

Il n'apparaît pas que l'utilisation quotidienne dans diverses institutions de l'outil informatique par les élèves sous la forme de logiciels fermés, prêts à l'emploi, ait pu profondément modifier les problèmes didactiques posés par la prescription à des élèves des années 2010 de tâches relevant de l'algorithmique-programmation, que ce soit pour faire des simulations informatiques d'expériences aléatoires, que ce soit dans le cadre d'un enseignement de l'algorithmique-programmation.

4. Les connaissances stables en algorithmique programmation

L'évolution rapide de la science informatique et du matériel ont pu amener dans les années 1980 à douter de la stabilité des connaissances dans cette discipline.

En 2010, la réalité de connaissances spécifiques à la science informatique n'est plus discutée.
Nous renvoyons à la thèse de Baron (1987) et aux manuels universitaires variés édités en 2011 pour accompagner un enseignement de l'algorithmique-programmation : trente ans d'écart n'ont pas fondamentalement changé les bases de l'algorithmique-programmation, même si certaines problématiques ont pu évoluer, en particulier les problèmes de complexité en temps et en espace, du fait du changement d'échelle des problèmes que l'informatique peut traiter, grâce essentiellement aux augmentations considérables de l'espace mémoire et de la rapidité de calcul.

Par contre les lignes de partages restent discutées entre les champs des mathématiques et de l'informatique.
Dans sa thèse, Nguyen Chi Thanh (2005) semble attribuer l'algorithmique aux

mathématiques pour ce qui concerne les algorithmes fondamentaux, les structures de données (tableaux, arbres, ensembles, listes, fichiers), l'analyse de complexité et les techniques de conception d'algorithmes :

« *Ainsi l'algorithmique cherche à étudier l'algorithme indépendamment de sa mise en œuvre dans une machine* ».

La programmation est alors selon Nguyen Chi Thanh « *l'ensemble des activités qui permettent l'écriture des programmes informatiques pour leur exécution dans une machine* ».

Il s'agit de transformer l'algorithme en un ensemble d'instructions dans un langage de programmation donné, instructions exécutables par la machine. Ces instructions sont structurées, dans les langages impératifs, par des instructions de contrôles : séquentialité, instructions conditionnelles, instructions itératives.

Sans s'enfermer dans la question de la légitimité de la place de l'algorithmique dans le champ des mathématiques (« programmes 2010 » de seconde et de première), on peut constater qu'il demeure beaucoup d'ambiguïtés : l'algorithmique étudie-t-elle l'algorithme (par exemple complexité en temps de calcul, complexité en espace mémoire) comme le disent Arsac (1988) précédemment cité, et Nguyen Chi Thanh, ou donne-t-elle des méthodes pour écrire des algorithmes, ou encore s'agit-il de quelque chose qui est très proche de la programmation, comme peut le laisser entendre l'utilisation du mot « algorithmique » dans certains contextes ?

Un document à destination des enseignants édité par le ministère de l'Education Nationale s'intitule : « *Algorithmique* » (document référencé [1] dans la bibliographie) ; on trouve dans sa tables des matières : « Quelques généralités sur l'algorithmique ». On s'attend à y trouver une définition du substantif « *l'algorithmique* », mais en fait, *algorithmique* retrouve dans ce paragraphe sont statut initial d'adjectif :

> *« La démarche algorithmique est, depuis les origines, une composante essentielle de l'activité mathématique. Au collège, les élèves ont rencontré des algorithmes (algorithmes opératoires, algorithme des différences, algorithme d'Euclide, algorithmes de construction en géométrie). Ce qui est proposé dans le programme est une formalisation en langage naturel. »*

Et plus loin, dans un sous-paragraphe intitulé « Algorithmes et démarche algorithmique » :

« La sensibilisation de l'élève à la question de la "démarche algorithmique" pourra se faire en évitant toute technicité ou exposé systématique. On pourra sur ce thème consulter des publications réalisées dans le cadre des IREM.
Les compétences suivantes pourront être identifiées et travaillées :
- comprendre et analyser un algorithme préexistant ;
- modifier un algorithme pour obtenir un résultat particulier ;
- analyser la situation : identifier les données d'entrée, de sortie, le traitement...;
- mettre au point une solution algorithmique : comment écrire un algorithme en "langage courant" en respectant un code, identifier les boucles, les tests, des opérations d'écriture, d'affichage... ;
- valider la solution algorithmique par des traces d'exécution et des jeux d'essais simples ;
- adapter l'algorithme aux contraintes du langage de programmation : identifier si nécessaire la nature des variables... ;
- valider un programme simple. »

L'acception ici du mot « algorithmique » est donc uniquement sous sa forme d'adjectif : « qui se rapporte aux algorithmes », et mis à part peut-être le quatrième point, elle prend en compte essentiellement ce qui se rapporte aux algorithmes déjà écrits (comprendre un algorithme préexistant, le modifier, identifier les données d'entrée, valider, adapter un algorithme).
Le quatrième point est assez ambigu : « *mettre au point une solution algorithmique* » ne veut pas dire la même chose que « trouver une solution algorithmique », et dans un contexte informatique, « mettre au point » renvoie plutôt à un travail sur le programme, c'est à dire sur l'algorithme codé dans un langage de programmation exécutable par la machine.
En fait, il semble que les « compétences » évoquées dans ce texte touchent de très près à la programmation.

La mise au point de l' « interface » entre un algorithme et le code (programme) de cet algorithme dans un langage de programmation donné est est une étape qu'on ne peut occulter : elle peut être extrêmement réduite pour les simulations d'expériences aléatoires que nous présentons ici, et beaucoup plus complexe pour l'écriture d'un programme pour résoudre une équation $f(x) = 0$ par la méthode de dichotomie à une précision donnée.

Lorsqu'on a pu encadrer la solution de l'équation à résoudre entre *a* et *b*, on calcule le signe de l'image de *(a + b)/2* par f, ce qui permet d'obtenir un nouvel encadrement de la solution, intervalle dont l'amplitude est la moitié de celle de l'encadrement précédent : savoir tout cela est nécessaire, et c'est pratiquement la définition de l'algorithme de résolution de l'équation par dichotomie. Cependant cela n'éclaire pas sur les lignes de programme qu'il convient d'écrire.

Il reste à trouver la « *façon de faire* » (Arsac (1988) déjà cité) ; ce n'est pas un problème d'algorithmique au sens mathématique : trouver une méthode mathématique pour résoudre un problème, étudier la complexité de cette méthode (éventuellement en prenant en compte son implémentation en machine) ; il s'agit « *seulement* [de] *trouver, face à un problème donné, une méthode qui en viendra à bout.* » (Arsac, idem).

Peut-être conviendrait-il de distinguer, au moins au niveau de l'enseignement secondaire, l'*algorithmique mathématique* de l'*algorithmique-programmation* : nous avons déjà employé cette formulation (avec ou sans trait d'union) pour laisser apparaître que nous notre propos se situait plus près du programme que de considérations de complexité (par exemple).

Cette question : « par quel bout commencer ? » renvoie à la mise en œuvre d'*heuristiques* au sens recommandé par la délégation générale à la langue française et aux langues de France (J.O. du 16 septembre 1989) : une heuristique est une *« méthode de résolution de problèmes, non fondée sur un modèle formel et qui n'aboutit pas nécessairement à une solution »*

Ce peut être par exemple la méthode des situations évoquée par Arsac (cf. paragraphe 2 de ce chapitre) pour construire une itération.

Ce peut être également une analyse descendante du problème, débouchant sur une décomposition du problème en problèmes plus simples, etc.

Dans l'introduction du cours polycopié intitulé « Algorithmes et programmation » pour les étudiants de l'école Polytechnique, Cori et Lévy (2011) écrivent :

> *« Notre cours doit être compris comme une initiation à l'informatique, et plus exactement à l'algorithmique et à la programmation. Il s'adresse à des personnes dont nous tenons pour acquises les connaissances mathématiques : nulle part on n'expliquera ce qu'est une récurrence, une relation d'équivalence ou une congruence. Il est orienté vers l'algorithmique, c'est-à-dire la conception*

d'algorithmes (et non leur analyse de performance), et la programmation, c'est-à-dire leur implantation pratique sur ordinateur (et non l'étude d'un - ou de plusieurs - langage(s) de programmations). (…)

Le cours est organisé selon les structures de données et les méthodes de programmation, utilisées dans différents algorithmes. Ainsi seront considérés successivement les tableaux, les listes, les piles, les arbres, les files de priorité, les graphes. En parallèle, on regardera différentes méthodes de programmation telles que la récursivité, les interfaces ou modules, la compilation séparée, le backtracking, la programmation dynamique. (…)

Le langage C a ses inconvénients, le principal est de ne pas avoir de garde-fous. On peut donc écrire des programmes inesthétiques, sous couvert d'efficacité. Pourtant, cet argument n'est même plus vrai avec l'avènement des machines RISC. On peut simplement écrire en C comme en Pascal. C'est le style que nous adopterons ici, et espérons que C pourra être considéré comme du "sucre syntaxique" autour de Pascal. »

Nous trouvons ici explicitement l'acception de l'algorithmique comme : la conception d'algorithmes.

La programmation d'un algorithme est conçue comme son « *implantation pratique sur ordinateur* ».

Cori et Lévy parlent de *style* de programmation : il y a plusieurs grandes classes de langages de programmation (langages impératifs, langages fonctionnels, langages déclaratifs, langages récursifs, tableurs etc., un même langage pouvant relever de plusieurs classes).

La mise en œuvre d'un algorithme dans une classe de langage donnée amènera un certain *style* de programmation.

Et comme le soulignent Cori et Lévy, à l'intérieur d'une même classe de langages, il peut y avoir divers *styles*. Ils donnent l'exemple de deux programmes en langage C pour calculer π :

« Ce programme fait un nombre inconsidéré d'appels au préprocesseur et beaucoup trop d'effets de bord. Il est illisible. Il faut avoir une conception simple et donc esthétique des programmes. On n'est plus à l'époque où tout se joue à l'instruction machine près. La puissance des ordinateurs a rendu très viable la programmation sans optimisations excessives. Tout l'art du programmeur, et notamment du programmeur C, est de comprendre les quelques endroits critiques

d'un programme, en général très peu nombreux, où les optimisations sont nécessaires. Donc le programme cryptique précédent, ou tout autre bourré d'optimisations inutiles sera mauvais et donc interdit. »

La différence entre les deux programmes évoqués – différence de style – relève d'un parti pris du programmeur (les optimisations), parti pris qui est extérieur à l'algorithme (il s'agit du calcul de π par le développement en série : $\pi/4 = arctan(1) = \sum (-1)^n / (2n+1)$), et qui est extérieur au langage de programmation.

5. Les *descriptions d'algorithme*

Pour ce qui concerne les simulations informatiques d'expériences aléatoires envisageables au lycée, il semble que l'on puisse laisser de côté la question : « comment construit-on un algorithme ? » dans la mesure où les algorithmes en jeu sont simples (*imitation au plus près* de l'expérience aléatoire), y compris pour ce qui concerne l'environnement informatique qui permettra d'effectuer un traitement statistique des résultats produits.

Par contre, il apparaît que la traduction d'un algorithme en un programme (code) exécutable sur machine, est conçue en tenant compte du langage de programmation dans lequel il doit être traduit (par exemple : le langage possède-t-il dans ses structures de contrôle les boucles « tant que », les boucle « jusqu'à » ?).
Nous faisons l'hypothèse que cette traduction pose pour les élèves plus de problème qu'il n'y paraît à lire les documents d'accompagnement des programmes proposés aux enseignants, documents que nous commenterons dans le chapitre suivant.

L'interface que nous évoquions dans le paragraphe précédent est une étape intermédiaire entre l'algorithme et le programme (programme au sens du code exécutable par la machine), étape que nous appellerons dans la suite : *description de l'algorithme*.
On n'envisage pas la même *description d'algorithme* selon que l'on s'autorise une programmation récursive ou non (la différence de classe de langages amène deux *styles de programmation* différents)
Ainsi pour la programmation du calcul de $n ! = 1 \times 2 \times ... \times n$, <u>en vue de la programmation dans un langage impératif</u>, la description de l'algorithme pourra être :

Entrer N

mettre F à 1

Pour J allant de 1 à N faire :

F reçoit F * J

Fin de faire

En vue de la programmation dans un langage récursif, la description de l'algorithme reposera sur la remarque que n ! = n × (n – 1) ! et sa description pourra se faire par :

Entrer N

Si N = 1 alors Factorielle(N) = 1 sinon Factorielle(N) = N * Factorielle(N-1)

La création d'une feuille de calcul sur tableur [61] et la création d'un programme dans un langage impératif ne peuvent pas a priori [62] reposer sur une même description d'algorithme :

• un programme dans un langage de programmation impératif est un enchaînement dans le temps d'instructions qui ont pour effet de modifier, au cours de l'avancement de l'exécution du programme, les valeurs qui se trouvent dans les diverses mémoires. L'interprétation (en général le sens, en relation avec les objectifs assignés au programme) de ces valeurs dépend du temps, c'est à dire du degré d'avancement de l'exécution du programme.

• l'interprétation des résultats d'une feuille de calcul de tableur est fondée sur les relations qu'entretiennent entre elles les cellules, une fois le calcul de la feuille effectué.

Cette étape de la *description de l'algorithme* produit un texte qui prend à la fois en compte la classe de langage de programmation dans lequel sera écrit le programme (code), et le *style de programmation* propre au programmeur.

Ce texte peut être traduit pratiquement directement dans le langage de programmation visé, et tout autre langage appartenant aux mêmes classes.

De tels textes seront présentés et commentés dans le prochain chapitre : ils sont extraits de

---

[61] nous laissons ouverte la question « peut-on encore parler de langage de programmation ? »

[62] même s'il est possible de « simuler » une boucle dont le nombre de répétitions est connu d'avance, comme cela est pratiqué précisément pour les simulations qui nous intéressent ici, lorsqu'elles se font sur tableur.

documents à destination des enseignants, accompagnant les « programmes 2010 ». Nous n'avons retenu que ceux qui ont trait aux simulations d'expériences aléatoires.

Ces textes sont présentés comme les « algorithmes », ce qui paraît impropre, tant ils sont pratiquement isomorphes à leur traduction dans le langage de programmation, et dans la mesure où l'on voit pour certains d'entre eux que le texte est conforme à certains partis pris, certaines heuristiques du programmeur, ce que nous avons appelé plus haut le *style de programmation du programmeur*. Nous préférons les considérer comme des *descriptions d'algorithmes* prêtes à être traduites dans les langages de programmation de la classe des langages pour laquelle ces descriptions ont été conçues.

CHAPITRE 11

Les simulations dans les documents d'accompagnement

Le Ministère de l'Education Nationale met à disposition des enseignants sur le site « éduscol » divers « *documents d'accompagnement »* et « *ressources pour la classe ».*

On peut trouver ainsi un document pour l'algorithmique [1], en relation avec les nouveaux programmes de mathématiques de seconde pour la rentrée 2009 (« programmes 2010 »), et un document pour l'enseignement des probabilités en première S (« Ressources pour la classe de première générale et technologique », « Statistiques et probabilités », [2], publié en juin 2011), documents dont nous donnons quelques extraits en annexe de ce chapitre, et dont on trouvera la référence à la fin de la bibliographie.

Nous allons analyser et commenter certains aspects de ces documents : le premier car il fait apparaître des problèmes didactiques liés à l'algorithmique-programmation, et le second car il ressort des différents exemples de simulation d'expériences aléatoires qui y sont présentés, que la tâche de simulation n'est pas analysée dans ce document, et que la finalité de la simulation n'y est pas toujours clairement définie.

1. *Descriptions d'algorithmes* et itérations

Dans la solution proposée par le premier document ([1]) pour réaliser la simulation d'une expérience aléatoire, les choix faits pour réaliser la simulation soulignent certains problèmes liés à une structure de donnée itérative ; ces problèmes paraissent mériter une grande attention de la part de l'enseignant qui voudrait mettre en œuvre cette simulation.

Il s'agit du « jeu du lièvre et de la tortue », déjà évoqué dans cette thèse, qui est abordé dans un paragraphe intitulé « Algorithmes et probabilités » ; la question posée est : *le jeu est-il à l'avantage du lièvre ou de la tortue ?* Le paragraphe commence par :

> « *Les algorithmes proposés ci-après s'insèrent dans le cadre des simulations, et par conséquent de l'approche dite "fréquentiste" des probabilités. »*

Nous reproduisons la suite du texte, en retrait et encadré ; nos commentaires sont intercalés :

a. Algorithme 1 : Simulation d'une partie sans boucle
La partie se finit en au plus six lancés. Il est donc possible de simuler une partie sans avoir recours à une boucle.

Variables
dé : la face du dé tirée au hasard
tour : compte le nombre de tours que dure la partie

Initialisation

dé **prend une valeur** entière aléatoire entre 1 et 6 compris
tour **prend la valeur** 1

Traitement
Si dé < 6 **alors**
dé **prend une valeur** entière aléatoire entre 1 et 6 compris
tour **augmente de** 1

Si dé < 6 **alors**
dé **prend une valeur** entière aléatoire entre 1 et 6 compris
tour **augmente de** 1

Si dé < 6 **alors**
dé **prend une valeur** entière aléatoire entre 1 et 6 compris
tour **augmente de** 1

Si dé < 6 **alors**
dé **prend une valeur** entière aléatoire entre 1 et 6 compris
tour **augmente de** 1

Si dé < 6 **alors**
dé **prend une valeur** entière aléatoire entre 1 et 6 compris
tour **augmente de** 1

Sortie
Si dé= 6 **alors**
Affiche « Le lièvre gagne »
sinon
Affiche « La tortue gagne »

Affiche tour

La variable tour paraît ici parfaitement inutile. Elle est peut-être le vestige d'un programme écrit avec une boucle, le test d'arrêt de la boucle s'appuyant sur la valeur de cette variable. Cette hypothèse illustrerait la remarque d'Arsac (1987) citée au paragraphe 2 du chapitre 10 : souvent dans l'enseignement, on écrit d'abord le programme, on le transpose en français, et on présente ensuite « la méthode de résolution du problème ».

Remarques :
– Si le support de programmation sur lequel est transposé l'algorithme ne dispose pas des fonctionnalités de recopie de texte, la mise en place d'une boucle peut se justifier.
– Il est possible de modifier un peu la modélisation du jeu, afin de simplifier sa mise en oeuvre, en particulier sur tableur ; en effet, la partie est équivalente aux lancers de six dés. Le lièvre gagne s'il existe au moins un six parmi les résultats.
– De cette façon un seul test peut suffire, à condition de disposer de l'instruction appropriée (comme l'instruction NB.SI du tableur).
– Cependant, l'arithmétique booléenne peut aussi se modéliser avec des sommes ou des produits d'entiers. Par exemple, le produit des six nombres (6 – *dé*) vaut 0 si et seulement il y a au moins un six.

La première remarque peut s'entendre comme une amorce de situation didactique pour introduire la notion de boucle.

La difficulté de l'équivalence d'une partie où l'on continue de jouer même si l'on sait que le lièvre a déjà gagné et d'une partie qui suit à la lettre la règle du jeu, n'est pas évoquée. Les « coups » inutiles ont historiquement posé problème, et ils en posent encore aux élèves.

La dernière remarque relève des « astuces de programmation » qui n'ont peut-être pas leur place dans une période d'initiation à l'algorithmique programmation.

b. Algorithme 2 : Cumuler un grand nombre d'expériences
Il suffit d'aménager le programme afin d'insérer la simulation d'une partie dans une boucle et de compter le nombre de parties gagnées par le lièvre ou la tortue. L'intérêt d'un langage de programmation devient évident : l'itération est très rapide aussi bien à écrire, modifier qu'à exécuter (ce qui n'est pas le cas avec le tableur).On pourra noter, à cette occasion, que certains langages sont beaucoup plus rapides que d'autres.

Variables
dé : la face du dé tirée au hasard
N : le nombre de parties à simuler
k : le compteur de boucle
tortue : le nombre de parties gagnées par la tortue
Initialisation
tortue **prend une valeur** 0
Traitement
Pour k **de** 1 **à** N
| dé **prend une valeur** entière aléatoire entre 1 et 6 compris
| **Si** dé < 6 **alors**
| | dé **prend une valeur** entière aléatoire entre 1 et 6 compris
|
| **Si** dé < 6 **alors**
| | dé **prend une valeur** entière aléatoire entre 1 et 6 compris
|
| **Si** dé < 6 **alors**
| | dé **prend une valeur** entière aléatoire entre 1 et 6 compris
|
| **Si** dé < 6 **alors**
| | dé **prend une valeur** entière aléatoire entre 1 et 6 compris
|
| **Si** dé < 6 **alors**

| | dé **prend une valeur** entière aléatoire entre 1 et 6 compris
| **Si** dé < 6 **alors**
| | tortue **prend la valeur** tortue + 1

Sortie
Affiche tortue .

La traduction proposée dans le langage XCAS est une traduction littérale de ce texte, nous ne la reproduisons pas.

c. Algorithme 3 : Avec une structure itérative conditionnelle.
Evidemment, plutôt que de répéter 6 fois les mêmes instructions, il est possible de simuler une partie à l'aide d'une boucle.
De cette façon, il sera facile d'expérimenter de nouveaux jeux en modifiant le nombre de cases que doit parcourir la tortue.
Variables
dé : la face du dé tirée au hasard
case : le numéro de la case sur laquelle se trouve la tortue
N : le nombre de cases que doit parcourir la tortue pour gagner.
Initialisation
N **prend la valeur** 6
case **prend la valeur 0.**
Traitement
Répète
| dé **prend une valeur** entière aléatoire entre 1 et 6 inclus.
| **Si** dé < 6 **alors**
| | case **prend la valeur** case + 1

jusqu'à [dé = 6 ou case = N **]**
Sortie
Si case = N **alors**
| **Affiche** « La tortue gagne »
Sinon
| **Affiche** « Le lièvre gagne »

Le choix a donc été ici de faire une boucle de type « jusqu'à » : on lance au moins une fois le dé (ce qui est bien le cas dans ce jeu), et après chaque lancer, on regarde si l'on doit s'arrêter ou continuer à lancer le dé. L'auteur de cette *description d'algorithme* semble avoir été guidé par une heuristique personnelle, peut-être liée à des considérations pédagogiques.

L'introduction de la variable N, initialisée à 6, afin, dit l'auteur, de pouvoir « expérimenter de nouveaux jeux en modifiant le nombre de cases que doit parcourir la tortue » fait se « téléscoper » la solution à deux problèmes ne natures tout à fait différentes : comment remplacer la répétition fastidieuse d'actions identiques (il suffit d'introduire une nouvelle structure de contrôle : les boucles), et comment s'offrir la liberté de changer de règle du jeu (on introduit là l'idée qu'un programme doit pouvoir être facilement modifié, ou adapté si le cahier des charges est modifié).

La détermination du test d'arrêt demande quelque réflexion, qu'il faudra détailler avec les élèves, en faisant un peu de logique : quand s'arrête-t-on ? Quand la partie est finie ! Et quand est-elle finie ? Quand l'un des deux adversaires a gagné, c'est à dire quand la tortue a gagné **ou** quand le lièvre a gagné. La tortue a gagné si case = N, le lièvre a gagné s'il sort un 6, d'où le booléen « dé = 6 ou case = N ».

Le problème est que l'auteur de la description d'algorithme a voulu traduire ce texte dans le langage SCILAB, qui n'admet pas les boucles « jusqu'à » :

Traduction SCILAB

```
N=6;
Ncase=0;
de=0;
while (de<6 | Ncase<N) do
de=floor(rand()*6+1);
if (de<6)
Ncase=Ncase+1;
end;
end;
if (Ncase==6)
disp("La tortue gagne");
else
disp("Le lièvre gagne");
end;
```

Remarques :
« *case* » est un mot-clé du langage SCILAB ; la variable s'appelle donc « *Ncase* ».

La structure *repeat..until* n'existe pas dans SCILAB, le code est donc légèrement aménagé par rapport à l'algorithme. Pour entrer dans la boucle une première fois, la variable « *de* » est initialisée avec la valeur arbitraire 0.

Nous avons fait ressortir en caractères gras le dernier paragraphe qui nous paraît mériter d'être commenté de façon détaillée.
Il ne s'agit en effet pas d'un léger aménagement du code par rapport à l'algorithme.
Pour un élève de seconde l'aménagement ne va probablement pas de soi.

D'autres heuristiques auraient pu amener une *description d'algorithme* d'emblée en termes de boucle « tant que » (nous préciserons dans la suite de ce paragraphe la description d'algorithme qu'on obtient alors), et parmi ces heuristiques pouvait figurer la prise en compte du langage dans lequel l'algorithme doit être implémenté.

L’aménagement suggéré ici consiste à transformer une boucle « jusqu’à » en une boucle « tant que », ce qui, pour un élève de seconde débutant en programmation, ne présente pas les mêmes enjeux qu’une *description d’algorithme* avec d’emblée une boucle « tant que » ; pour réaliser la transformation , il faut :

• comprendre que le test d’arrêt est la négation du booléen précédent, et être capable d’effectuer cette négation,

• comprendre le problème du premier passage nécessaire dans la boucle qui amène à initialiser la variable *dé* « *avec la valeur arbitraire 0* ».

Cette valeur n’est pas tout à fait arbitraire, puisqu’il faut initialiser *dé* avec un entier strictement inférieur à 6. Pourquoi avoir choisi 0 pour une variable qui représente une face d’un dé cubique ? Du point de vue du **sens** que l’on attribue aux variables informatiques, ce choix n’est peut être pas judicieux. En fait dans cette « re-traduction », la variable « dé » a deux fonctions, ce qui peut prêter à confusion chez des élèves débutants.

Reprenons donc la *description de l’algorithme* avec en perspective une boucle « tant que » :
On répète tant que ; tant que quoi ? Tant que ce n’est pas fini…
Envisageons donc une variable booléenne « fini », qui au départ est à la valeur *faux*
Que doit-on faire ? Lancer le dé et récupérer sa valeur dans X
Si X = 6, le lièvre a gagné et c’est fini : la variable « fini » doit passer à *vrai*.
Sinon, la tortue avance d’un pas et il faut regarder si elle est arrivée au but ; il faut donc un compteur de pas « Pas » ; au départ (avant le premier lancer de dé, donc avant l’entrée dans la boucle), « Pas » doit être à zéro.
Si « Pas » = 6, la tortue a gagné, et « fini » passe à vrai.
Récapitulons :

fini prend la valeur *faux*
Pas prend la valeur 0.
Tant que fini = *faux* faire
    X prend une valeur entière aléatoire entre 1 et 6 inclus
    Si X = 6 alors afficher « le lièvre gagne » et fini prend la valeur *vrai*
        sinon augmenter Pas de 1
    FinSi
    Si Pas = 6 alors afficher « la tortue gagne » et fini prend la valeur *vrai*

FinSi

FinTantque

Une autre heuristique (par exemple, un souci pédagogique tel que faire de la logique avec des élèves ?) peut amener à approfondir le booléen « fini » :

Quand est-ce fini ? Si l'un des deux adversaires a gagné, c'est à dire si

X = 6 ou « Pas » = 6 ; donc quand n'est-ce pas fini ?

Si NON[X = 6 **ou** « Pas » = 6], c'est à dire :

si NON(X= 6) **et** NON(« Pas » = 6), c'est à dire :

si X < 6 **et** « Pas » < 6 :

D'où la *description d'algorithme* (provisoire) :

Pas prend la valeur 0

Tant que (X < 6 et Pas < 6) faire

X prend une valeur entière aléatoire entre 1 et 6 inclus

Si X= 6 alors afficher « le lièvre gagne »

sinon augmenter Pas de 1

FinSi

Si Pas = 6 alors afficher « la tortue gagne »

FinSi

FinTantque

Mais il manque l'initialisation de X qui autorisera le passage au moins une fois dans la boucle ; avec des élèves débutants, il est probablement préférable d'aborder le problème de front : le booléen **(X < 6 et Pas < N)** doit être à *vrai* au départ ce qui autorise toute valeur de X inférieure strictement à 6.

On voit sur cet exemple que *les descriptions d'algorithmes* ne sont pas toutes équivalentes, que ce soit pour la traduction dans le code d'un langage de programmation donné, que ce soit pour les problèmes didactiques qui risquent d'être soulevés.

Du point de vue *algorithmique mathématique*, l'algorithme est de faible complexité : une boucle de longueur au plus six, quelle que soit la *description d'algorithme* choisie.

2. La tâche de simulation dans les « programmes 2010 »

Les exemples donnés dans les documents d'accompagnement de première S ([2] « Ressources pour la classe de première générale et technologique, Statistiques et probabilités », juin 2011), confirment une technique pour la tâche de simulation semblable à celle que nous avons décrite dans le chapitre 8, et que nous avons qualifiée d'*imitation au plus près*.

2.1. L'aspect algorithmique de la tâche

Nous remarquons que dans ce document qui est destiné aux professeurs, aucun commentaire concernant l'aspect algorithmique-programmation n'accompagne les exemples de simulation traités : ni pour le choix du style de programmation (lié au choix du langage de programmation, ici essentiellement le tableur ou un langage impératif), ni pour les difficultés de programmation possibles, ni pour l'interprétation des résultats.

Ainsi le premier exemple (page 5) concerne une « *approche heuristique* » (l'expression figure dans les « programmes 2010 » de première) de la loi des grands nombres à partir de l'observation de simulations (X est égale au numéro sorti lorsqu'on lance un dé équilibré) :

> *« A l'aide d'une simulation, on répète un grand nombre de fois cette expérience aléatoire à l'identique et on peut ainsi observer un grand nombre de réalisations de la variable aléatoire X. Le graphique suivant montre l'évolution de la moyenne observée en fonction du nombre de répétitions. »*

Le graphique présenté semble provenir d'un tableur. La taille de l'échantillon est 1000. Il n'est pas fait mention du choix de l'échelle du graphique qui semble être le choix automatique du tableur : l'échelle sur l'axe des ordonnées est choisie à partir de l'amplitude des données. Nous avons montré (chapitre 2, paragraphe 5) que ce choix masque la différence entre la « tendance à se rapprocher de », convergence en probabilité, et la convergence d'une suite numérique.

Le deuxième exemple de simulation, page 14, concerne la loi géométrique tronquée : on répète dans des conditions identiques une expérience de Bernouilli de paramètre p avec au maximum n répétitions, et arrêt du processus au premier succès ; la loi géométrique tronquée de paramètre n et p est la loi de la variable aléatoire X qui est nulle si aucun succès n'a été obtenu, et égale à k si le premier succès obtenu l'a été à l'étape k.

Notons que le terme « indépendants » pour qualifier les essais est remplacé par « conditions identiques ».

Le document propose une simulation de cette loi. Le choix est fait, sans justification, d'un langage impératif. Il est vrai que la condition d'arrêt du processus, qui est : « tant qu'il n'y a pas eu de succès et que le nombre d'essais ne dépasse pas *n* », évoque une boucle « tant que », ce qui ne se programme pas aisément sur tableur.

Cependant, le nombre d'essais étant de toute façon limité par n, il est possible, au risque d'avoir des « essais pour rien », de programmer cette loi sur tableur. Nous avons évoqué les problèmes posés par les essais pour rien, et on peut penser qu'il est préférable de les éviter.
Mais si c'est l'élève qui doit faire la simulation, on ne peut évacuer trop vite l'étape du *style de programmation* à choisir pour concevoir cette simulation : un élève peut être renté de faire cette simulation sur tableur si on ne lui impose pas un langage impératif. Et si on lui impose d'éviter le tableur, il parait souhaitable d'expliquer pourquoi (ou de faire découvrir pourquoi).
L'enseignement de l'algorithmique programmation est, dans ces « programmes 2010 » conçu au travers d' « activités transversales », telles les simulations, pour lesquelles la question du *style de programmation* et du temps de calcul ne peut être éludée.

Le document que nous étudions ici propose ce que nous appelons une *description d'algorithme* pour langage impératif, puis traduit cette description pour une calculatrice « TI » (« Texas Instrument »), pour « Algobox » et pour « Scilab ». Les programmes obtenus sont certes isomorphes, mais n'auront pas la même efficacité du point de vue du temps de calcul : il convient, avant de démarrer l'activité de simulation, de se poser la question des objectifs de cette simulation : s'il s'agit de faire un exercice d'algorithmique-programmation dans un langage impératif, la loi géométrique tronquée n'étant qu'un prétexte à une boucle « tant que », la programmation sur une calculatrice a le mérite de pouvoir être menée en dehors d'une salle informatique. S'll s'agit de faire une étude statistique d'un grand nombre de simulations pour diverses valeurs de *n* (page 52 de ce document), la calculatrice est inadaptée.

2.2. Les hypothèses de modélisation

Les « programmes 2000 » insistaient fortement sur le choix du modèle, un modèle parmi d'autres possibles. Nous avons montré que ce choix est en fait le choix des *hypothèses de modélisation*, choix qui va déterminer aussi bien la loi de probabilité que l'on veut associer à l'expérience, que sa simulation.
Dans les exemples présentés dans ce document, il n'est pas souligné, lorsqu'une simulation est proposée, où et comment sont traitées les hypothèses de modélisation ; l'accent n'est donc

pas mis sur le lien entre le modèle probabiliste et la simulation de l'expérience (quel que soit l'ordre dans lequel l'un et l'autre sont présentés), lien qui s'établit précisément au travers d'un choix identique d'hypothèses de modélisation.

On peut même remarquer que les hypothèses de modélisations ne sont peut-être pas présentées avec la rigueur que l'on pourrait attendre. Ainsi :

• « *On considère que chaque enfant a une chance sur deux d'être un garçon ou une fille* » (page 19). On peut penser qu'après sa naissance, le sexe de l'enfant ne relève plus de l'aléatoire…

• Dans le paradoxe de Saint-Petersbourg (cf. le document 8 annexé à ce chapitre), l'hypothèse que la pièce est équilibrée est absente.

• A propos d'une association (page 35), il est écrit : « *Les statistiques montrent que chaque adhérent assiste à l'assemblée avec la probabilité 80%* ». Il est difficile d'imaginer quelle étude statistique a permis d'assigner à un adhérent la probabilité de 0,8 d'assister à l'assemblée. S'il s'agit d'une étude à partir de la proportion d'adhérents ayant assisté à la précédente assemblée, par exemple 80 %, le problème posé (quelle est la probabilité que le quorum de 50 % soit atteint lors de la prochaine assemblée générale ?) perd tout intérêt : si l'on considère que la proportion de participants est stable, 80 % de l'effectif participera à la nouvelle assemblée, donc le quorum sera atteint.

La solution attendue (mettant en œuvre la loi binomiale de paramètres 0,8 et 30) donne une probabilité de 0,99994761.

On peut remarquer avec Fine (2010) qu'il y a parfois des raccourcis fâcheux qui amènent à confondre la probabilité de tirer, dans une population où la proportion d'un certain caractère est p, un individu ayant ce caractère, avec la probabilité qu'un individu ait ce caractère.

2.3. Le statut des simulations dans le document [2]

Nous avons observé comment les simulations d'expériences aléatoires étaient présentées dans l'ensemble de ce document « Ressources pour la classe de première générale et technologique, Statistiques et probabilités », : l'objectif (déclaré ou non) de la simulation, ou sa mise en page (avant ou après le modèle probabiliste de l'expérience, etc.).

• Il y a parfois confusion entre l'expérience et sa simulation, c'est le cas par exemple pour la simulation produisant des résultats dont l'observation est « une approche heuristique » de la loi des grands nombres (pages 4 à 6) : « *on répète un grand nombre de fois cette expérience aléatoire »*, mais en réalité, c'est sa simulation qu'on répète. L'expérience simulée (le lancer du dé) n'est qu'un prétexte à observer une autre expérience aléatoire (sa simulation), que l'on peut répéter un très grand nombre de fois.

• Concernant la loi géométrique tronquée, la présentation qu'en fait le document est la suivante :

- ❖ *Définition de la loi géométrique tronquée*
- ❖ *Expression de la loi géométrique tronquée*
- ❖ *Algorithme de simulation*
- ❖ *Représentation graphique*
- ❖ *Espérance de la loi géométrique tronquée*

La détermination du modèle probabiliste précède la simulation (qui est une *imitation au plus près*).

Les représentations graphiques sont faites à partir de la simulation, et permettent d'énoncer des conjectures qui sont alors démontrées.

• Un exemple d'activité est proposé, lié à la loi géométrique tronquée : la limitation des naissances, d'après Claudine Robert (2003) :

Enoncé

Pour limiter le nombre de filles dans un pays (imaginaire ?), on décide que :

chaque famille aura au maximum 4 enfants ;

chaque famille arrêtera de procréer après la naissance d'un garçon.

On considère que chaque enfant a une chance sur deux d'être un garçon ou une fille et que, pour chaque couple de parents, le sexe d'un enfant est indépendant du sexe des précédents.

Ce choix a-t-il la conséquence attendue, à savoir de diminuer le nombre de filles dans la population ?

Simulation de l'expérience sur un tableur

Représentation à l'aide d'un arbre

La simulation précède ici le calcul de la loi de probabilité associée aux issues (qui prennent en compte le nombre total N d'enfants et le nombre total G de garçons) puis de l'espérance des variables N et G, et enfin le calcul de E(G)/E(N).

Il est dit : « *La simulation montre clairement que la proportion de garçons semble bien rester voisine de 0,5. La politique nataliste mise en place n'aurait donc aucun effet sur la modification de cette proportion.* » La simulation a été faite sur tableur (avec essais inutiles pour résoudre la difficulté d'une boucle « tant que » sur tableur), et pour 1000 familles la proportion observée de garçons est de 0,495.

Aucun commentaire n'est fait sur les essais inutiles, ni sur la fluctuation d'échantillonnage. Ce seul résultat de 0,495 fait apparaître une proportion inférieure à 0,5, qui pourrait être interprétée comme une baisse de la proportion de garçons.

• Le paradoxe de Saint-Petersbourg est un autre exemple d'activité liée à la loi géométrique tronquée (page 22) :

Enoncé

Un joueur joue contre la banque au jeu de "pile ou face", en misant toujours sur "face". Il adopte la stratégie suivante : il mise un euro au premier coup, et s'il perd, double la mise au coup suivant, tant que "face" ne sort pas. S'il gagne, il récupère sa mise augmentée d'une somme équivalente à cette mise. Le joueur dispose d'une fortune limitée, qui lui permet de perdre au maximum n coups consécutifs et, si "pile" sort n fois de suite, le joueur ne peut plus miser et arrête le jeu. La fortune de la banque, elle, n'est pas limitée.

Une partie consiste pour le joueur à jouer, si sa fortune le permet, jusqu'à ce que "face" sorte.

Il s'agit de déterminer la probabilité qu'a le joueur de gagner une partie, son gain algébrique moyen par partie, et d'analyser l'intérêt pour le joueur de jouer à ce jeu.

Traitement mathématique

Simulation de 1000 parties en 9 coups au plus sur un tableur

Conclusion de l'étude : deux paradoxes

Le choix des sous-titres pourrait laisser entendre que la simulation n'est pas un « traitement mathématique », alors que ce paragraphe renvoie à une étude statistique des résultats de la simulation.

Là encore, la simulation est faite sur tableur, avec des coups pour rien.

Le premier paradoxe souligné par les commentaires est que le jeu semble favorable au joueur puisque avec une fortune illimitée la probabilité qu'il gagne une partie tend vers 1, donc sa stratégie « *constitue une "martingale" infaillible* », alors que « *puisque E(Y) = 0, le jeu est honnête* ». Mais la fortune du joueur n'est pas illimitée…

La simulation est utilisée alors pour montrer les limites du calcul de l'espérance pour juger si un jeu est favorable au joueur : la répétition, par exemple de 1000 simulations fait bien ressortir que si la probabilité de perdre une partie est faible, la perte est importante, alors que le gain est faible, même si le joueur gagne souvent ; lors des 1000 simulations observées, le joueur a perdu 6 fois 511 euros et a gagné 994 fois un euro ; le joueur a donc perdu 2972 euros sur 1000 parties : c'est le deuxième paradoxe annoncé par le document, qui conclut :

> *« La notion de risque, liée à celle de la dispersion de la variable aléatoire "gain", est un élément décisif d'appréciation d'un jeu. Le paradoxe de Saint-Petersbourg est l'un des problèmes ayant donné naissance à la théorie de la décision en économie. Dans cette théorie, on formalise en particulier la notion de* ***fonction d'utilité****, qui mesure le degré de satisfaction d'un consommateur . »*

L'ouverture à la théorie de la décision est intéressante, mais ne renvoie pas à des références bibliographiques. Cependant, sans aller jusqu'à cette théorie, cette simulation offre la possibilité de développements vers la notion de risque, développements que n'appelait pas, a priori, la détermination de la seule loi de probabilité, et que nous évoquerons dans la quatrième partie.

TROISIEME PARTIE

Du côté des élèves : expérimentation avec quatre classes

Nous allons dans cette partie rendre compte d'une expérimentation que nous avons menée à la fin de l'année scolaire 2008-2009 et pendant l'année scolaire 2009-2010 dans des classes de Première et Terminale S.

Nous souhaitions observer le comportement d'élèves en situation de classe pendant des activités de simulation et de modélisation.
Plus précisément, nous avons observé les productions écrites par chacun de ces élèves pour répondre à un travail demandé (précisé sur un énoncé imprimé) à faire en classe.
Le travail demandé consistait à déterminer la loi de probabilité d'une expérience aléatoire, à construire une feuille de calcul pour simuler cette expérience, et à proposer une méthode pour résoudre un problème lié à cette expérience aléatoire.

Cette expérimentation s'inscrit dans les objectifs que nous nous étions assignés lorsque nous avons débuté cette thèse : tenter d'éclairer la tâche de simulation, son lien avec la tâche de modélisation, et son éventuel apport pour une maîtrise des notions de probabilité dans le cadre de l'enseignement en classes de Première et Terminale de lycée, aux fins éventuelles de créer une ingénierie didactique.
C'est pourquoi nous avons tenu à privilégier le cadre de la classe.

CHAPITRE 12

1. Les objectifs de notre expérimentation

Deux devoirs, donnés à faire en classe à des élèves de première S afin d'évaluer la compréhension du lien entre la valeur moyenne observée et l'espérance mathématique (lien qui figure au programme de première S), nous ont permis de commencer à vérifier notre hypothèse selon laquelle on peut attendre dans l'enseignement secondaire deux types de simulation d'une expérience aléatoire :

- une *imitation au plus près* des différentes actions en jeu dans l'expérience aléatoire, et alors la tâche est essentiellement de type algorithmique ;
- une programmation de la loi de probabilité, et la tâche relève essentiellement des probabilités.

Le travail proposé aux élèves lors de notre expérimentation s'inspire en partie de l'un de ces devoirs, qui a pu nous servir de pré-expérimentation.

1.1. Mettre les élèves en situation de créer une simulation : une pratique récente

Mettre les élèves en situation de créer eux-mêmes une simulation d'une expérience aléatoire est une pratique récente, sur laquelle la communauté enseignante n'a pas encore dégagé de consensus pédagogique.

1.1.1. Une pratique récente

La nécessité d'intégrer au cours de probabilité l'activité de création de simulations par les élèves eux-mêmes n'était pas explicitement énoncée dans les « programmes 2000 » [63] ; elle s'est imposée avec les premiers sujets d'épreuve pratique expérimentale dans l'année scolaire 2006-2007, puisque cette épreuve devait passer du statut d'expérimental au statut d'épreuve obligatoire à la session de juin 2009.

Il n'y a pas eu, à notre connaissance, d'explications sur les motivations ou justifications de ce

---

[63] voir les programmes de Première S et Terminale S, en annexes de la première partie. On remarquera le vague des formulations «(…) on illustrera ceci par des simulations dans des cas simples » ou « On simulera des lois de probabilité simples (…) ». Quel est l'acteur qui se cache derrière ce « on » : l'enseignant ou l'élève ?
Le style habituel des programmes (et les autres paragraphes de **ces** programmes) donnent la réponse : « on » est l'enseignant.

changement d'acteur : c'est l'élève qui doit créer des simulations.

Il n'y a pas eu de formation particulière en direction des enseignants pour cette tâche de simulation d'expériences aléatoires : ni pour les aider à réaliser eux mêmes cette tâche, ni pour les préparer à prendre en charge ce qu'il faut bien appeler un enseignement de cette tâche.

1.1.2. Pas de consensus pédagogique, semble-t-il, sur l' « intérêt » de faire des simulations

Zaki (1990) qui a observé l'activité d'étudiants ayant à résoudre trois problèmes de probabilité en confrontant les résultats de leurs calculs avec les résultats de simulations (il attendait des étudiants interrogés qu'ils fassent des estimations à partir des résultats observés) conclut :

> *« La pratique de la simulation a donc eu quelques effets favorables sur le développement du jugement probabiliste des étudiants. Cependant, nous serons plus réservés quant au rôle didactique que peut jouer la simulation dans l'enseignement des probabilités : le seul fait de procéder à la réalisation d'expériences aléatoires ne semble pas être suffisant (comme on a pu le constater chez la plupart des étudiants) pour améliorer les démarches probabilistes des étudiants, et encore moins pour les amener à développer des traitements inductifs. En revanche, on peut identifier par la pratique de la simulation certaines conceptions probabilistes erronées (...), et profiter de cette occasion pour apporter aux étudiants des explications plus convaincantes, où la simulation d'expériences aléatoires va servir de support d'interprétation. »*

Bordier (1991) se fixait pour objectif, au début de ses recherches, de corriger certaines conceptions erronées de la probabilité. Il constate à propos d'une recherche qu'il a menée auprès de sujets jouant régulièrement au *Loto 6/49* :

> *« Malgré les résultats probants de cette étude, au plan de la correction des conceptions erronées, de tels environnements de simulation offrent assez peu d'intérêt pour l'enseignement de la probabilité. L'utilisateur n'arrive pas à saisir, de cette façon, pourquoi ce qu'il croyait vrai auparavant, s'avère faux dans les faits. (…) Un logiciel de simulation peut éventuellement modifier des croyances et des habitudes, mais il ne permet pas, à lui seul, une compréhension des*

phénomènes en jeu. »

Bordier, dans cette étude, avait utilisé un logiciel dédié à ce jeu de loto. Il estime préférable, plutôt que d'utiliser un logiciel dédié à la simulation d'une situation aléatoire particulière, d'utiliser un « laboratoire » :

> *« S'il* [le laboratoire] *a, selon nous, un grand potentiel éducatif, c'est qu'il crée un cadre à l'intérieur duquel la tâche principale de l'élève n'est pas d'observer une simulation déjà construite, mais plutôt d'en définir qui soient bien adaptées à l'étude de tel ou tel phénomène aléatoire ou encore à la résolution d'un problème. Dans ce contexte, plus ouvert, la part de l'élève est beaucoup plus importante que s'il était en présence d'un logiciel déjà constitué. Mais la situation didactique ainsi créée semble favoriser le type de réflexion requis pour faire évoluer les conceptions des élèves sur la probabilité. »*

Wozniak (2005) craint que le recours aux simulations des expériences aléatoires ne soient un prétexte pour enseigner la statistique et les probabilités en restant dans des mathématiques « épurées », éloignées du monde réel.

En dehors des thèses de Zaki (1990), Bordier (1991) et Coutinho (2001), nous n'avons pas trouvé d'étude récente comportant des observations d'élèves de lycée, afin d'étudier l'apport des simulations (informatiques ou non) d'expériences aléatoires dans l'apprentissage de la statistique et des probabilités.

Dagnelie (2010) semble avoir constaté chez ses étudiants une tendance à considérer que l'on ne peut pas faire une étude statistique avec un seul échantillon, idée qui serait confortée par l'habitude prise de multiplier les échantillons lorsque les étudiants travaillent avec des simulations informatiques.
Roditi (2009) utilise des simulations sur tableur pour introduire le test du χ^2 auprès d'étudiants en sciences humaine et sociales. La feuille de calcul est préparée par l'enseignant.

Dans la noosphère, après dix ans de mise en œuvre des « programmes 2000 », les objectifs assignés à l'utilisation de simulations informatiques d'expériences aléatoires sont assez divers, et parfois contradictoires. Voici un rapide tour d'horizon d'objectifs de travaux menés (ou à mener) avec des élèves, objectifs que avons pu recueillir sur les site des IREM à l'été

2010 :

- valider le modèle probabiliste par des simulations,
- valider la simulation par le modèle probabiliste (déterminé après la simulation de l'expérience aléatoire),
- accéder à la notion de modèle par des simulations (cf. l'article de Parzysz (2009) commenté au chapitre 4).
- faire, avec des simulations, une approche fréquentiste de la loi de probabilité d'une expérience aléatoire
- introduire des tests d'hypothèse par des simulations.

1.2. Pas de travaux, semble-t-il, sur la nouvelle tâche dévolue aux élèves

Il n'y a pas eu, à notre connaissance, de recherche didactique sur la tâche de simulation d'expérience aléatoire maintenant dévolue aux élèves.
Cette tâche apparaît dans les « programmes 2010 » à la fois sous l'angle « exercice d'algorithmique-programmation », nous en avons donné un aperçu au chapitre 11, et sous l'angle « illustration de notions de probabilités » (ou introduction à des notions de probabilité) : nous avons analysé de ce point de vue des documents d'accompagnement de la classe de première S au chapitre 11.

Rappelons que ces documents d'accompagnement des « programmes 2010 » de la classe de première S, destinés aux enseignants, présentent des exemples résolus de simulations informatiques d'expériences aléatoires.
Nous avons montré que ces exemples ne sont pas commentés du point de vue algorithmique-programmation : le choix du style de programmation n'est pas justifié, et ne paraît pas toujours approprié ; la finalité de la simulation, dont nous avons montré qu'elle peut influencer et le style de programmation et le choix du matériel sur lequel le programme sera implémenté, n'est pas définie au départ comme partie intégrante du *cahier des charges*.
Définir cette finalité est également un moment important de l'étude probabiliste du problème, de même que le choix des hypothèses de modélisation sur les expériences élémentaires composant l'expérience que l'on veut simuler.
Dans ces documents d'accompagnement (« programmes 2010 »), les hypothèses de modélisation ne sont pas toujours explicitées, et elles le sont parfois de façon assez relâchées.
Selon le modèle que nous avons présenté au chapitre 9 des liens entre l'expérience, son modèle probabiliste et sa simulation, ces hypothèses de modélisation sont précisément le lien

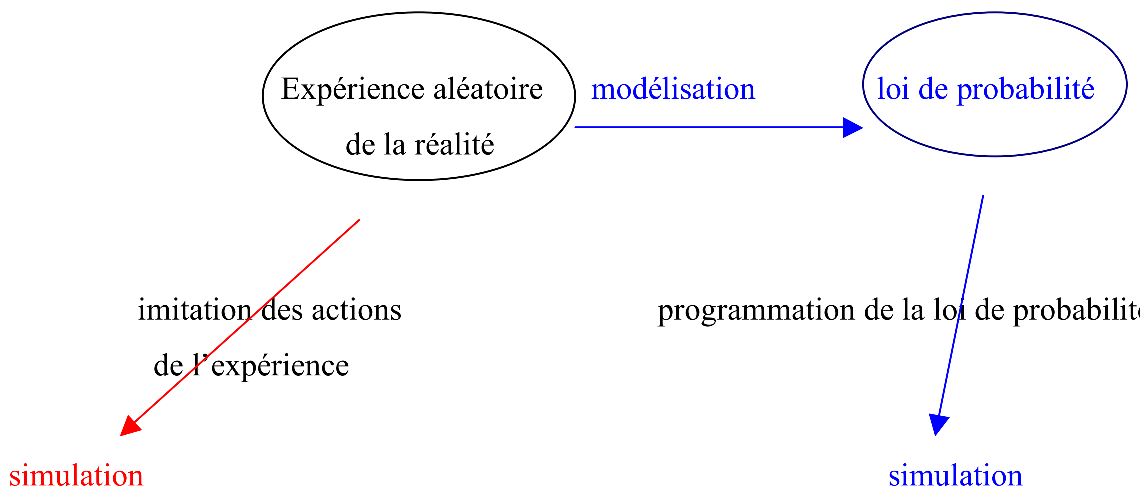
entre le modèle probabiliste et la simulation de l'expérience aléatoire étudiée.
Ce document ne fait à aucun moment apparaître ce lien, et ne suggère pas à l'enseignant d'être attentif aux procédures mises en œuvre par les élèves lorsqu'ils simulent une expérience aléatoire.

Nous avons voulu observer le comportement d'élèves de lycée ayant, entre autres, à construire une simulation, et avons organisé une expérimentation autour de deux axes d'observation.

2. Le premier axe d'observation

Plus précisément, nous avons voulu observer les procédures mises en œuvre par les élèves pour simuler une expérience aléatoire composée, pour laquelle il est possible de voir quel choix d'hypothèses de modélisation est fait dans la constitution de la simulation et dans la détermination de la loi de probabilité, et comment ces hypothèses sont mises en œuvre.

Nous souhaitions également observer quelles techniques seraient mises en œuvre : une *imitation au plus près* de l'expérience aléatoire, la simulation de sa loi de probabilité, ou d'autres techniques ?
Le schéma 1 qui suit décrit deux techniques possibles pour construire une simulation d'une expérience aléatoire.
Schéma 1



Ce que nous entendons ici par « simulation » est un programme informatique (dans notre expérimentation : une feuille de calcul), dont l'exécution est la réalisation d'une expérience aléatoire, qui a même modèle que l' « expérience aléatoire de la réalité », qui lui est équivalent, si la simulation est correcte.

Notons que les deux chemins n'amènent pas le même programme informatique (dans notre expérimentation la même feuille de calcul), même si nous considérons l'un et l'autre des programmes obtenus comme deux simulations correctes de l'expérience aléatoire initiale : leurs exécutions fournissent deux expériences aléatoires équivalentes à l'expérience aléatoire, puisqu'ayant même modèle probabiliste (cf. chapitre 6).
Nous pensons utile de préciser l'un et l'autre des chemins indiqués sur le schéma 1.

Pour le chemin rouge, l' « imitation des actions de l'expérience » est une activité de type algorithmique, dont les points forts sont :
• la définition du *cahier des charges*,
• sa traduction en une description d'algorithme adaptée au type de programmation visé : tableur, ou langage impératif, ou langage récursif, par exemple,
• la programmation de cette description d'algorithme.
Dans la pratique, ces points ne sont pas toujours clairement dissociés ; les recommandations méthodologiques qui ont pu être prodiguées autrefois aux élèves de l'option informatique insistaient fortement sur l'ordre de ces étapes, qu'il est impératif de respecter. L'enseignement de la discipline informatique en lycée dans les années 1980 n'était peut-être pas exempt de dogmatisme sur ce point : il s'agissait de réagir à des pratiques non rigoureuses, tant dans les milieux professionnels que dans les clubs informatiques des lycées.

Pour le chemin bleu, « programmer » la loi de probabilité, cela peut être simuler le tirage d'une boule d'une urne dont la composition est convenablement choisie.
Cette technique suppose que les probabilités des n issues telles qu'elles ont été déterminées dans le modèle probabiliste de l'expérience aléatoire initiale, soient des nombres rationnels $\frac{p_k}{q}$ ($k = 1, 2, \ldots, n$) où les p_k et q sont des entiers, et $\sum p_k = q$: l'urne contient alors q boules possédant k qualités différentes observables (par exemple k couleurs différentes), l'urne contenant p_k boules ayant la $k^{ème}$ caractéristique. Réaliser l'issue k avec la probabilité p_k revient à tirer de l'urne une boule ayant la caractéristique k.

Schéma 2

La réalisation pratique de la simulation d'un tel tirage ne pose pas de difficulté particulière : on tire un nombre entier aléatoire x entre 1 et q selon une loi équirépartie ; si $1 \leq x \leq p_1$ la boule a la première couleur ; si $p_1 + 1 \leq x \leq p_1 + p_2$, la boule a la deuxième couleur, etc.

Lorsque la loi de probabilité a été déterminée par transport d'une loi équirépartie par une variable aléatoire, les probabilités des issues associées à la variable aléatoire ont été obtenues par la formule $\frac{\text{nombre de cas favorable}}{\text{nombre de cas possible}}$, et les probabilités des issues seront des nombres rationnels.

Il peut se trouver que les probabilités soient présentées sous forme décimale : il est certes toujours possible de les ramener à des fractions d'entiers ; il est également possible de simuler une expérience de même loi de probabilité sans passer par un tirage d'urne, particulièrement si le logiciel de programmation ne possède qu'une fonction aléatoire « Alea » donnant des nombres (pseudo) aléatoires sur [0 ; 1] selon la loi uniforme. Ainsi pour une expérience de Bernouilli à deux issues, avec la probabilité p d'avoir un succès :

si Alea < p alors Résultat = « succès » sinon Résultat = « échec »

On aura le résultat « succès » avec une probabilité p.

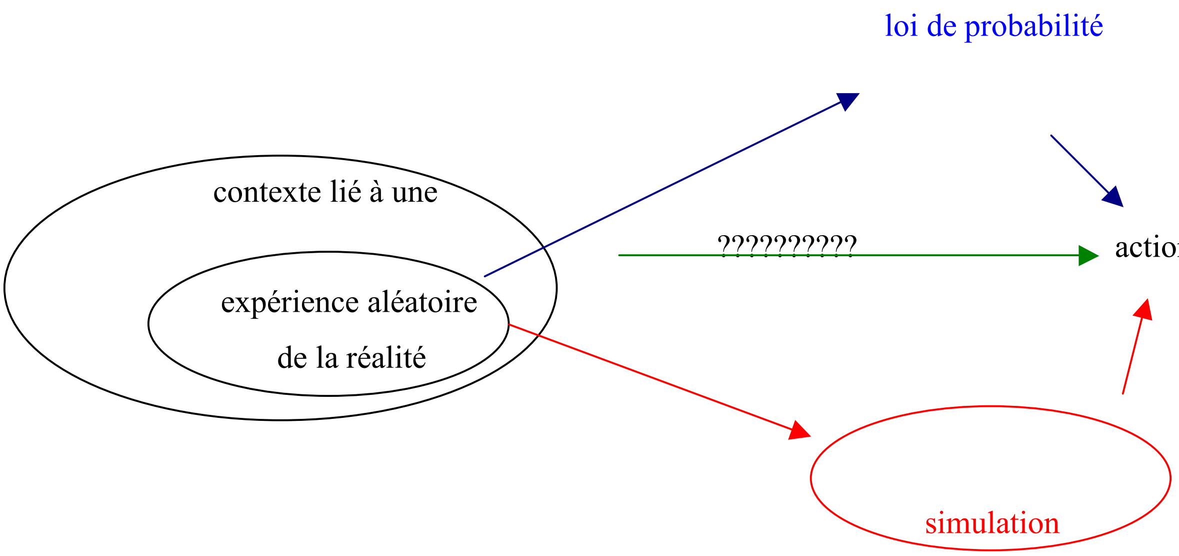
Pour une expérience aléatoire ayant plus de deux issues, on procède comme dans la simulation d'un tirage d'urne par un ensemble de tests sur le tirage obtenu par « Alea ».

La technique de simulation de l'expérience aléatoire par simulation de sa loi de probabilité sera alors mieux représentée par le schéma 1.
En effet, la simulation/programmation de la loi de probabilité s'est faite sans le détour par la représentation de la loi de probabilité par un tirage d'urne.
Il n'est pas sûr que le détour par un tirage d'urne soit indispensable, ni judicieux. Ce détour n'est peut-être qu'un vestige des simulations d'expériences aléatoires à l'aide des didacticiels fermés que nous avons évoqués, pour lesquels les tirages d'urnes étaient un passage obligé.

3. Le deuxième axe d'observation

Lorsqu'un élève dispose de la loi de probabilité et de la simulation d'une expérience aléatoire, et qu'il doit agir dans un contexte lié à cette expérience, va-t-il s'appuyer sur ces éléments ? Si oui, comment ?

<u>Schéma 3</u>



Nous avons représenté, en bleu, en rouge et en vert trois voies possibles, pas nécessairement exclusives. La voie en vert peut aussi bien représenter une absence d'action.

Il nous appartiendra, dans l'étude des réponses des élèves, d'essayer d'analyser aussi finement que possible les diverses procédures développées sur chacune de ces voies, tout particulièrement la référence éventuelle à la loi des grands nombres, ou le calcul dans le modèle probabiliste d'indicateurs tels que l'espérance mathématique d'une variable aléatoire.

CHAPITRE 13

Présentation du travail demandé aux élèves

Nous avons proposé un travail autour d'une expérience aléatoire composée, dont l'élève devait à la fois établir la loi de probabilité et une simulation.
Nous avons organisé ce travail selon deux modalités : la modélisation probabiliste de l'expérience précédant sa simulation, ou la simulation précédant la modélisation probabiliste.

Les deux tâches sont explicitement demandées sur l'énoncé distribué aux élèves. Elles précèdent une question ouverte liée à cette expérience aléatoire.

1. Les énoncés

La partie en italique est commune à tous les élèves ; elle est suivie de questions, selon deux modalités (Enoncé 1 et énoncé 2 : l'ordre des questions A.1. et A.2. est inversé)

> *Monsieur Dupond veut inventer un nouveau jeu de plateau, pour jouer à deux ou trois joueurs.*
>
> *Dans ce jeu, le joueur dont c'est le tour lance deux dés tétraédriques équilibrés. Il avance sur le plateau d'un nombre de cases égal à l'écart entre les points amenés par les deux dés (dans le cas d'un double, le joueur n'avance pas). Le gagnant est le premier joueur à atteindre la case d'arrivée.*
>
> *On a conseillé à Monsieur Dupond de limiter la durée du jeu à 30 minutes.*
> *Monsieur Dupond évalue à 1,5 minutes la durée nécessaire pour que se joue un tour avec trois joueurs (c'est à dire pour que l'un après l'autre chacun des trois joueurs lance les dés et avance sur le plateau selon le résultat obtenu).*
> *Monsieur Dupond calcule donc qu'il faudrait que l'un des joueurs gagne après au plus 20 tours ($1{,}5 \times 20 = 30$ minutes).*
>
> *Monsieur Dupond se demande alors combien de cases il doit y avoir entre la case de départ et la case d'arrivée pour pouvoir* ***espérer*** *qu'il y aura un gagnant dans les délais impartis, c'est à dire à l'issue de 20 tours au plus.*

Enoncé 1 :

A. 1. Lorsqu'un joueur joue à ce jeu, il avance de X cases. X peut prendre les valeurs 0, 1, 2 et 3. Complétez le tableau page 2 donnant la probabilité de chacune des valeurs possibles de X. Vous expliquerez <u>en détail</u> votre raisonnement.

2. On souhaite simuler avec un tableur une partie de 20 tours pour un joueur. Indiquez sur la feuille de calcul représentée en page 2 les formules que vous placeriez dans les cellules pour faire cette simulation et connaître le nombre de cases parcourues par ce joueur.

B. Comment conseilleriez-vous Monsieur Dupond pour l'aider à prendre sa décision concernant le nombre de cases séparant la case de départ de la case d'arrivée ?

Enoncé 2 :

A. 1. On souhaite simuler avec un tableur une partie de 20 tours pour un joueur. Indiquez sur la feuille de calcul représentée en page 2 les formules que vous placeriez dans les cellules pour faire cette simulation et connaître le nombre de cases parcourues par ce joueur.

2. Lorsqu'un joueur joue à ce jeu, il avance de X cases. X peut prendre les valeurs 0, 1, 2 et 3. Complétez le tableau page 2 donnant la probabilité de chacune des valeurs possibles de X. Vous expliquerez <u>en détail</u> votre raisonnement.

B. idem sujet A

La page suivante donne un cadre dans lequel l'élève doit répondre aux questions. Nous reproduisons ici ce cadre, sans respecter la mise en page de ce qui fut distribué aux élèves.
Bien entendu les champs pour les réponses aux questions A.1. et A.2. sont inversés dans les énoncés 1 et 2 :

A. 1.

X =	0	1	2	3
probabilité correspondante				

Justifications :

A. 2. Feuille de calcul du tableur

	A	B	C	D
1				
2				
20				
21				
22				

B. Comment conseiller Monsieur Dupond ?

La dernière page est un mémento des commandes du tableur qui peuvent être utiles :

> Rappels des principales commandes d'Open Office, et de leur syntaxe
>
> • SI(test;valeur-si-vrai;valeur-si-faux)
>
> • par exemple SOMME(A1:A100) calcule la somme de toutes les cellules de A1 jusqu'à A100
>
> • MOYENNE(A1:A100) calcule la moyenne de toutes les cellules de A1 jusqu'à A100
>
> • les commandes statistiques MAX, MIN, MEDIANE et ECARTYPE suivent la même syntaxe ; QUARTILE(A1: A100;1) calcule le premier quartile de la série qui se trouve dans les cellules de A1 à A100 ; si on remplace le 1 par un 3, on obtient le troisième quartile
>
> • ALEA.ENTRE.BORNES(n;p) sort un nombre entier aléatoire entre les entiers n et p, selon une loi équirépartie
>
> • les fonctions EXP, LN, LOG, ABS, ENT [64], SIN, COS, etc. *fonctionnent de manière classique*
>
> • les connecteurs logiques ET et OU *fonctionnent en notation préfixée* :
>
> ET(valeur logique 1; valeur logique 2 ; ....) renvoie "VRAI" si tous les arguments sont vrais, "FAUX" sinon
>
> OU(valeur logique 1; valeur logique 2; ....) renvoie "VRAI" si au moins un argument est vrai, "FAUX" sinon.

2. Quelques remarques sur les énoncés

Nous allons faire quelques remarques sur les énoncés du problème, plus particulièrement sur le choix de certains mots et l'ordre des questions et le rôle de Monsieur Dupond.

2.1. Le mot « espérer »

Nous avons choisi de faire ressortir en caractères gras le mot « *espérer* » pour souligner le caractère incertain du contexte, prenant peut-être ainsi le risque d'orienter, dans la partie B.,

[64] ABS(B3) : valeur absolue du nombre qui se trouve en B3 ; ENT(B4) : partie entière du nombre qui se trouve en B4.

la réflexion des élèves vers un calcul d'espérance mathématique (notion du programme de première S). Cependant ce mot n'apparaît pas dans le corps des questions.

2.2. Une question ouverte

La partie B. correspond à notre deuxième axe d'observation : observer si les élèves utilisent le modèle probabiliste et/ou la simulation déterminés dans la partie A. Nous ne pensions pas trouver des solutions complètes au problème, mais espérions que les élèves s'impliqueraient suffisamment dans la recherche d'une solution pour laisser des indices sur la manière dont ils ont intégré (ou n'ont pas intégré) ce qu'on peut faire avec une simulation ou une modélisation d'une situation aléatoire.

Nous souhaitions donc que les élèves lisent la question B. comme une question ouverte au sens précisé par l'Inspection Générale de mathématiques : les sujets de baccalauréat peuvent depuis la session de septembre 2008 présenter des questions ouvertes, signalées comme telles aux candidats ; une note [65] aux concepteurs de sujets de baccalauréat précise :

> *Exemple de texte porté dans l'en-tête du sujet :*
> *« Le candidat est invité à faire figurer sur la copie toute trace de recherche, même incomplète ou non fructueuse, qu'il aura développée. Il est rappelé que la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies. »*
>
> *Exemple de texte inséré avant une question nécessitant une prise d'initiative :*
> *« Dans cette question [dans cet exercice], toute trace de recherche, même incomplète, ou d'initiative même non fructueuse, sera prise en compte dans l'évaluation. »*

Non seulement le candidat sait alors qu'il a à résoudre une question ouverte (c'est à dire sans indication sur une méthode de résolution possible), mais il sait également que l'on n'attend pas nécessairement de lui une résolution complète de la question, mais qu'il présente un ou plusieurs « angles d'attaque » qu'il aura essayés pour résoudre cette question.

Nous n'avons pas voulu dans notre énoncé reprendre l'une de ces formulations, ne souhaitant pas que les élèves (particulièrement de Terminale, familiers avec les sujets de baccalauréat) se sentent dans des conditions d'examen. Nous avons voulu cependant, par le choix de la question (« comment conseiller », et non « que conseiller ») indiquer qu'était attendue la

[65] Au 25 août 2011, cette note a disparu du site « Eduscol » et du site de l'Inspection Générale de Mathématiques ; nous la reproduisons dans son intégralité en annexe de ce chapitre.

présentation d'une démarche, plutôt qu'un résultat numérique ; nous avons pris l'assurance auprès des enseignantes des classes concernées que chacune des classes a rencontré avant l'expérimentation des questions ouvertes.

2.3. La présentation des questions

La structure de l'énoncé nous a semblé importante, en particulier la numérotation des questions.

Il est fréquemment indiqué dans les énoncés du baccalauréat que pour faire une question, le candidat peut admettre les résultats des questions précédentes, et la règle est indiquée dès la classe de seconde.

Les enseignants, dès la seconde, et particulièrement en terminale, exercent leurs élèves à exploiter, dans une première lecture suivie d'un énoncé, la numérotation des questions, pour deviner des enchaînements logiques, en sachant néanmoins que deux questions consécutives peuvent être tout à fait indépendantes.

Les présentations :

1.	ou	A.1
2.		A.2
3.		B.

ne sont en général pas équivalentes : la présentation de gauche laisse plutôt pressentir une progression entre les 3 questions, une nécessité peut-être d'utiliser le résultat d'une question pour faire la suivante. La présentation de droite peut laisser entendre que la partie B. est d'une certaine manière déconnectée de la partie A. Cependant il y a un lien entre les deux parties (sinon elles seraient présentées comme des exercices indépendants), et l'ordre a peut-être un sens :

• A. peut être un cas particulier, et B. une généralisation ;
• A. peut contenir un ensemble de résultats dont certains peuvent être utiles pour B. ;
• A. peut présenter une étude générale, et B. un cas particulier.

Cette question de la présentation des question renvoie à l'influence possible de la formulation de l'énoncé du problème dans la façon dont les élèves vont tenter de le résoudre (Maury (1985) et (1986)).

La numérotation en deux parties A. et B. était susceptible de permettre à l'élève d'interpréter la question B. comme une question ouverte où il est possible d'investir des résultats de la partie A. , alors qu'une numérotation 1., 2., 3. l'aurait peut-être plutôt amené à chercher une réponse à la question 3. utilisant la question 2, privilégiant ainsi, selon la modalité de l'énoncé, la simulation ou la modélisation probabiliste pour résoudre la question posée dans la partie B.

Nous souhaitions par ailleurs que l'ordre des questions dans la partie A. soit une variable de commande de notre expérimentation (simulation avant ou après la détermination de la loi de probabilité), afin d'observer si cette variable a des effets
- sur la technique utilisée pour simuler l'expérience aléatoire dans la partie A
- sur la démarche imaginée pour « conseiller M. Dupond » (la proximité de la question A2 peut orienter vers une démarche s'appuyant sur ce qui a été déterminé en A2 (la loi de probabilité ou la simulation) plutôt que ce qui a été déterminé en A1.

Première modalité : énoncé 1	Deuxième modalité : énoncé 2
A.1. Programmez une simulation A.2. Déterminez la loi de probabilité B. Comment conseiller M. Dupond ?	A.1. Déterminez la loi de probabilité A.2. Programmez une simulation B. Comment conseiller M. Dupond ?

2.4. Monsieur Dupond

Ce problème est susceptible d'introduire la notion de risque : le risque que la partie ne se termine pas dans les délais impartis. C'est donc un risque que doit assumer le concepteur du jeu (Monsieur Dupond), qui va peut-être commercialiser ce jeu, et qui aura à assumer toutes les conséquences de ses choix.
Le « conseilleur » (l'élève, un groupe d'élève, la classe) a la responsabilité de souligner l'existence de ce risque, d'essayer de le mesurer, afin de permettre à Monsieur Dupond de choisir le risque qu'il veut bien assumer.
Nous avons voulu, en introduisant ce personnage de M. Dupond, permettre à l'élève de comprendre le risque en termes de risque qu'il doit essayer d'évaluer de façon objective en fonction des conditions expérimentales du jeu, et non en fonction du risque que lui, élève, est prêt à assumer.
Cette question nous a semblé être importante pour une exploitation possible de ce problème

par les enseignantes avec les classes, ultérieurement à l'expérimentation.

3. Une pré-expérimentation

Un exercice donné dans un devoir en classe de première S, l'analyse des réponses des élèves que nous avons été amenée à faire, ainsi que les prolongements qui ont accompagné la correction de ce devoir en classe (en particulier un devoir à la maison), ont pu nous servir de pré-expérimentation, particulièrement pour la partie A. (loi de probabilité et simulation), car le contexte de cet exercice était très voisin de celui du travail donné dans notre expérimentation ; en fait c'est l'analyse des copies des élèves pour ce devoir qui nous a suggéré le travail que nous avons proposé aux élèves dans notre expérimentation.

On veut inventer un nouveau jeu de plateau. Dans ce jeu, le joueur dont c'est le tour lancera deux dés tétraédriques équilibrés. Il avancera sur le plateau d'un nombre de cases égal à l'écart entre les points amenés par les deux dés (dans le cas d'un double, le joueur n'avance pas). Il est prévu que ce jeu s'arrêtera au plus tard après 25 tours. Le gagnant sera le premier joueur à atteindre la case d'arrivée.

1°) Lorsqu'un joueur joue à ce jeu, il avance de X cases. X peut prendre les valeurs 0, 1, 2 et 3. Recopiez et complétez le tableau ci-dessous donnant la probabilité de chacune des valeurs possibles de X. Vous expliquerez <u>*en détail*</u> *votre raisonnement.*

X =	***0***	***1***	***2***	***3***
probabilité correspondante				

2°) On se demande quel doit être le nombre de cases à prévoir pour espérer qu'il y aura un gagnant à l'issue des 25 tours. Pour cela, on décide de simuler 25 tours pour un joueur.

a. *Indiquez sur la feuille annexe les formules que vous placeriez dans les cellules du tableur pour faire cette simulation.*

b. *Comment utiliseriez-vous cette simulation pour décider du nombre de cases ?*

Nous avions pu observer, à l'occasion de ce devoir en classe, qu'une majorité d'élèves avait proposé une simulation par *imitation au plus près*, aucun élève ne proposant une simulation de la loi de probabilité qu'il venait pourtant d'établir. Nous avions également observé, pour quelques élèves, une discordance entre les hypothèses de modélisation qui semblaient avoir été choisies pour la loi de probabilité et pour la simulation : loi équirépartie sur {1 ; 2 ; 3 ; 4} pour chacune des deux expériences élémentaires que sont le lancer de chacun des deux dés, ou bien loi équirépartie sur l'ensemble {0 ; 1 ; 2 ; 3} des quatre valeurs que peut prendre la variable aléatoire « nombre de cases à avancer ». Nous avions de plus observé que le choix erroné avait pu être fait pour la simulation ou pour la modélisation probabiliste.

Le même devoir en classe nous a servi dans une moindre mesure de pré-expérimentation pour la question ouverte de notre expérimentation ; si le problème posé était le même, à savoir déterminer le nombre de cases à mettre sur le jeu pour que la partie soit terminée à l'issue d'un nombre donné de tours (25), dans ce devoir, la méthode était suggérée aux élèves dans l'énoncé : utiliser la simulation de 25 tours qui venait d'être conçue pour un joueur ; par ailleurs, le paramètre « nombre de joueurs » apparaissait très en retrait dans la formulation de l'énoncé.
L'objectif dans cette dernière question du devoir était d'observer le nombre de simulations jugées utiles ou nécessaires par l'élève pour pouvoir exploiter cette simulation, et le lien fait éventuellement par l'élève entre la moyenne des valeurs observées pour le nombre X de cases avancées sur 25 tours, et l'espérance mathématique de cette variable X.

Cette question, que nous avions conçue comme une question ouverte pour les élèves, l'était également pour nous en ce sens que nous n'avions encore jamais tenté d'évaluer ces points, qui figurent cependant dans le « programme 2000 » de première S : nous avons rencontré des réponses très variées (de l'absence de réponse, à une prise de décision basée sur *un très grand nombre de simulations* : valeur moyenne ou plus petite valeur observée, en passant par une prise de décision fondée sur <u>une</u> simulation) ; certaines réponses signalaient l'importance du nombre de joueurs, d'autres essayaient d'exprimer le lien entre la valeur moyenne observée et l'espérance mathématique d'une variable aléatoire, d'autres essayaient de se donner une marge de manœuvre en diminuant la valeur moyenne observée de 5 cases. Nous transcrivons ici des extraits de trois copies, <u>sans les commenter</u> :

> « *Il faudra donc relancer la simulation un très très grand nombre de fois et faire la moyenne des cases avancées après 25 tours. Comme la simulation de ces 25 tours aura été faite un très grand nombre de fois, on peut être presque sûr que la moyenne des cases avancées conviendra car d'après la loi des grands nombres, plus n est grand (n étant ici le nombre de simulations) plus on se rapprochera de la valeur théorique (ici la valeur théorique est la moyenne des cases avancées après n lancers).* »

Dans une autre copie :

> « *(...) et on lui* [la valeur moyenne] *soustrait un nombre assez grand pour espérer qu'il y ait un gagnant à chaque partie, par exemple 5, on obtient donc le nombre de cases à prévoir pour le jeu.* »

Et enfin :

> « *Il faut faire attention à la façon dont on peut exploiter les informations obtenues grâce à la simulation : je pense que plutôt que de se baser sur des simulations reproduisant des cas réels possibles, il est préférable de calculer le nombre de cases minimum pour être sûr d'avoir fini avant le 25 ème tour.*
> *Cependant, si je veux absolument utiliser les résultats de la simulation, je la répèterai autant de fois que possible (Ctrl Alt F9) puis mon nombre de cases sera le plus petit nombre que j'ai vu apparaître dans la case C26.* »

La variété des propositions des élèves pour la dernière question de ce devoir nous a amenée à envisager un travail autour du même problème, mais avec une question plus ouverte encore, c'est à dire n'orientant pas a priori vers l'exploitation de la simulations, et faisant par ailleurs plus nettement apparaître le nombre de joueurs comme un paramètre du problème.

CHAPITRE 14

Analyse a priori

Notre analyse a priori, concernant la modélisation et la simulation demandées aux élèves, est fondée d'une part sur les solutions des questions, telles que nous les envisagions lors de la conception du travail à donner aux élèves, et d'autre part sur les observations que nous avions pu faire à partir du devoir en classe et de ses prolongements évoqués au chapitre précédent, qui nous ont servi en quelque sorte de pré-expérimentation. Dans ce qui suit, c'est à ce devoir et ses prolongements que nous nous référons lorsque nous parlons de « pré-expérimentation ».

1. Analyse a priori de la modélisation demandée

Les deux dés sont équilibrés. Les issues peuvent être représentées par les cases d'un tableau à deux entrées, chacune des entrées étant associée au numéro de la face sortie sur chacun des deux dés. Cette représentation des 16 issues par une case de ce tableau est une représentation présentée de manière classique, en classe de première, pour déterminer la loi de probabilité de la variable aléatoire égale à la somme des points amenés lorsqu'on lance deux dés cubiques équilibrés :

dé n° 1 / dé n° 2	1	2	3	4	5	6
1		3				
2	3					
3						
4						
5						
6						

Chaque case vide représente donc une issue ; du fait du choix des hypothèses de modélisation (les deux dés sont équilibrés), une case n'a pas plus de chance de sortir qu'une autre : la loi va être équirépartie sur l'ensemble des 36 issues. Il ne reste plus qu'à inscrire la somme des points amenés correspondant à chacune des issues, et à appliquer le théorème pour une loi équirépartie : $p(A) = \dfrac{\text{nombre de cas favorables à A}}{\text{nombres de cas possibles}}$; ainsi, p(S=3) = 2/36.

Cet exercice est un exercice de base en première S, souvent proposé en illustration de la loi de probabilité d'une variable aléatoire lors de l'introduction de la notion de variable aléatoire. Il est souvent prolongé par la détermination de la loi de probabilité d'une autre variable aléatoire définie sur le même univers, telle que le plus grand des numéros amenés, ou l'écart absolu des numéros amenés.

Avec deux dés tétraédriques, cette représentation des issues, et l'utilisation des cases pour inscrire l'écart absolu des points amenés (en rouge) amène le tableau :

dé n° 1 / dé n° 2	1	2	3	4
1	0	1	2	3
2	1	0	1	2
3	2	1	0	1
4	3	2	1	0

On en déduit alors la loi de probabilité cherchée du nombre X de cases avancées par un joueur sur un tour lorsqu'il lance les deux dés tétraédriques équilibrés, loi de probabilité présentée de façon courante dans un tableau :

$x_i =$	0	1	2	3	total
$p(X = x_i)$	4/16	6/16	4/16	2/16	1

tableau 1

Cette solution est la solution que l'on peut attendre particulièrement d'élèves de première S.

Cependant on introduit en première S également, pour certaines expériences aléatoires, une représentation des issues de l'expérience avec des objets mathématiques rencontrés dans d'autres contextes, tels que des couples, triplets, ou ensembles : (x ; y), (x ; y ; z), {x ; y}, {x ; y ; z}, que l'on peut être amenés à généraliser : n-uplets $(x_1 ; x_2 ; \dots ; x_n)$ ou $\{(A_1 ;a_1) ; \dots (A_n ; a_n)\}$, si cette généralisation n'a pas déjà été fait dans le « chapitre statistique » ou lors de l'introduction de familles de n points pondérés, à l'occasion du « chapitre barycentre ».

Dans ce cadre là, une représentation de l'ensemble des issues peut être :

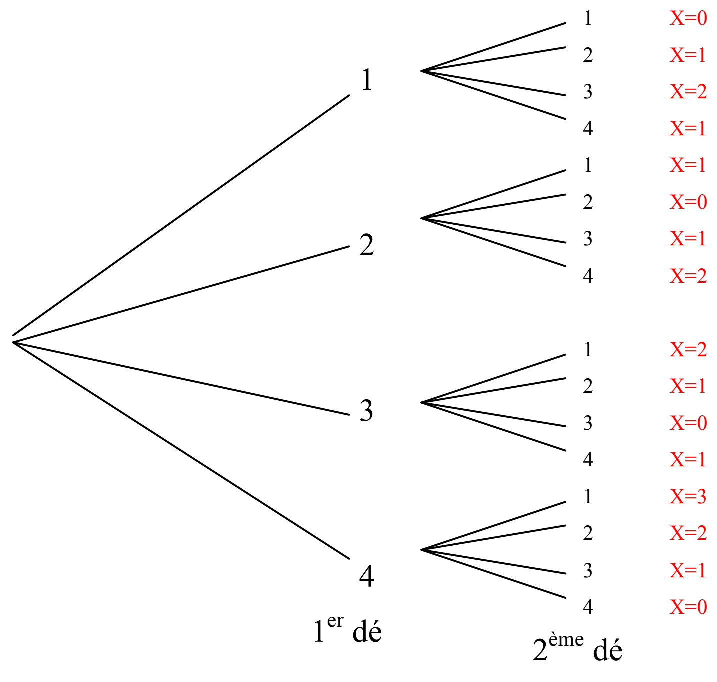
$\Omega = \{(x ;y) / x \in \{1 ; 2 ; 3 ; 4\} \text{ et } y \in \{1 ; 2 ; 3 ; 4\}\}$ pour une définition en compréhension de l'univers,

ou : Ω ={(1 ; 1) ; (1 ;2) ; ….. ; (4 ; 4)} pour une définition en extension, complètement écrite ou seulement suggérée.

L'hypothèse que les deux dés sont équilibrés amène à exprimer que les couples ont la même probabilité.

Le dénombrement des issues ainsi représentées par des couples est un exercice classique de dénombrement, qui peut être « soutenu » par une représentation arborescente de ces couples. Le dénombrement des *issues favorables* est moins immédiat : il peut consister à construire, pour chacune des valeurs de X, l'ensemble des issues favorables ; ainsi : X=0 pour x = y, ce qui donne 4 couples, X=1 pour (2 ; 1), (1 ; 2), (3 ; 2), (2 ; 3), (4 ; 3) et (3 ; 4), ce qui donne 6 couples, etc.

Le dénombrement des issues favorables peut aussi s'appuyer sur un arbre qui représente les issues ; les hypothèses de modélisation vont se traduire par : puisque les dés sont équilibrés, toutes les feuilles de l'arbre sont équiprobables. On peut mettre en regard de chaque feuille la valeur de X pour cette issue :



Certains élèves veulent représenter chacun des deux dés, manifestant par là qu'ils ont reconnu une expérience composée, mais représentent les issues de l'expérience par des **paires** : soit en les énonçant « un 1 et un 2 », soit en écrivant de manière plus formelle {1 ; 2} ; ils traduisent les hypothèse de modélisation « les deux dés sont équilibrés » par l'équiprobabilité des issues ainsi décrites. Nous avons rencontré cette faute lors de la pré-expérimentation et nous la rencontrons régulièrement dans nos classes de la part d'élèves débutants, avec le lancer de deux dés cubiques.

La loi de probabilité alors proposée est :

$x_i =$	0	1	2	3	total
$p(X = x_i)$	4/10	3/10	2/10	1/10	1

tableau 2

Nous avons également rencontré, lors de la pré-expérimentation, des solutions où l'élève prend pour univers l'ensemble {0 ; 1 ; 2 ; 3} qui est l'univers image de l'univers à 16 éléments par la variable aléatoire X, c'est à dire l'univers des *issues intéressantes* pour l'expérience (cf. chapitre 7, paragraphe 9).

Ces solutions occultent l'aspect composé de l'expérience aléatoire en considérant d'emblée, comme issues de l'expérience, les *événements intéressants* : avancer de 0 case, avancer d'une case, etc. ; ces solutions associent la loi équirépartie à ces 4 issues, et illustrent le biais d'équiréprobabilité.

Remarquons à nouveau que le passage des hypothèses de modélisation : « les deux dés sont équilibrés », c'est à dire <u>deux</u> hypothèses, à une loi équirépartie sur un univers à 36 éléments ne peut être détaillé par les élèves. En effet, faute d'une modélisation du lancer de deux dés en termes d'espace produit des deux espaces probabilisés modélisant le lancer de chacun des deux dés, ce passage ne peut se faire que dans les termes « puisque les dés sont équilibrés je n'ai pas plus de chances de sortir sur chacun d'eux un 1 qu'un 2, etc., donc je n'ai pas plus de chance de sortir avec les deux dés le couple (1 ; 3) que le couple (2 ; 3) , donc les issues de mon univers Ω sont équiréparties ».

Cela fonctionne correctement avec les couples ; la phrase reste pertinente si l'on remplace les couples (1 ; 3) et (2 ; 3) par les paires {1 ; 3} et {2 ; 3} , ou par l'énoncé « je n'ai pas plus de chances de sortir un 1 et un 3 qu'un 2 et un 3 », mais la conclusion sur l'équirépartition des issues représentées par des paires est alors fausse.

Lorsqu'un élève énonce directement l'affirmation : « les dés sont équilibrés donc la loi est équirépartie sur l'ensemble des issues » (ensemble correct à 16 éléments, ou ensemble erroné à 10 éléments), il est difficile de savoir si l'équiprobabilité des issues est bien le résultat d'un raisonnement tel que celui que nous venons de décrire, qui distingue les hypothèses de modélisation (les deux dés sont équilibrées), de la conclusion d'équirépartition des issues

physiques.
En revanche l'erreur qui consiste à proposer la loi équirépartie sur l'ensemble {0 ; 1 ; 2 ; 3} relèver du biais d'équiprobabilité : nous ne voyons pas de lien possible entre cette loi uniforme sur {0 ; 1 ; 2 ; 3} et les hypothèses de modélisation sur les deux dés.

Cette erreur se distingue également de l'erreur où l'univers, constitué de paires, a 10 éléments, en ce qu'elle occulte comme nous l'avons déjà dit le caractère composé de l'expérience : on va directement aux *événements intéressants* dont on fait les issues de l'expérience, mais en oubliant le détour par les *issues physiques*, liées aux expériences élémentaires qui composent l'expérience étudiée, détour qui justement permet de transporter, comme disent les programmes, la loi de probabilité sur les *issues physiques*, en une loi sur des issues plus intéressantes.

Dans l'hypothèse, que nous allons retenir, où l'équiprobabilité sur les issues d'un univers à 16 ou 10 éléments n'est pas une manifestation du biais d'équiprobabilité, les deux lois de probabilité (l'une juste et l'autre fausse) présentées dans les **tableaux 1 et 2** des pages précédentes, relèvent d'une mise en œuvre correcte des hypothèses de modélisation sur les dés ; l'erreur qui consiste à choisir un univers à 10 éléments est une description insuffisante des *issues physiques* sur lesquelles portent les hypothèses de modélisation : l'élève a oublié de mettre de la couleur sur les dés pour les distinguer.

2. Analyse a priori de la simulation demandée

Nous avons rencontré, dans notre pré-expérimentation, majoritairement, une *imitation au plus près* de l'expérience, et n'avons jamais rencontré la simulation de la loi de probabilité, qui avait cependant été déterminée à la question précédente ; des élèves n'ont pas proposé de simulation.
Nous avons également rencontré la simulation de la loi équirépartie sur l'ensemble {0 ;1 ; 2 ; 3}, erreur relevant du biais d'équiprobabilité, comme dans le paragraphe précédent lorsque la loi uniforme est proposée comme loi de probabilité pour la variable X.

Le plus souvent la feuille de calcul, dans la pré-expérimentation, a été organisée de la manière suivante : une ligne par tour (le tableau que nous avions fourni y incitait fortement, le nombre de colonnes étant faible et les lignes étant numérotées de 1 à 28) ; sur chaque ligne : une cellule pour simuler chacun des dés, puis calcul de l'écart par différentes procédures, le plus souvent par l'écriture mathématique de la valeur absolue entre deux barres, comme indiqué en bleu dans le tableau ci-dessous (les élèves n'avaient pas rencontré la fonction valeur absolu sur tableur, et nous n'avions pas distribué de mémento des principales commandes du tableur).

Nous avons représenté ci-dessous le tableau donné à remplir aux élèves, avec dans les cellules les formules proposées par une majorité d'élèves (en couleur) ; les flèches indiquent une recopie des formules vers le bas.

L'année de la pré-expérimentation, les élèves avaient travaillé sur un tableur, dans une version qui proposait la seule fonction aléatoire ALEA() qui rend un nombre (pseudo) aléatoire dans l'intervalle [0 ; 1] selon la loi uniforme. La formule pour obtenir un entier entre 1 et 6 selon la loi équirépartie est expliquée par l'enseignant lors de la première simulation d'un dé cubique équilibré (c'est un exercice suggéré dans le programme de statistique de seconde) et semble bien retenue par les élèves. Tous les élèves qui ont répondu à la question ont redonné cette formule lors de la pré-expérimentation.

	A	B	C	D	E
1	n° du lancer	Premier dé	Deuxième dé	écart	
2	1	=ENT(ALEA()*4+1)	=ENT(ALEA()*4+1)	=|B2 - C2|	
3	2	↓	↓	↓	
	…				
26	25				
27					
28			Somme des points amenés	=SOMME(D2 :D26)	

En dehors de la « simulation » fausse signalée ci-dessus (biais d'équirépartition), et de quelques copies où la question n'a pas été traitée, presque toutes les solutions rencontrées lors de la pré-expérimentation sont très voisines de celle qui est présentée dans ce tableau ; les

variations portent sur la cellule choisie pour faire calculer la somme des nombres de cases avancées par tour (par exemple en E2 ou E3), ou sur les formules proposées pour calculer l'écart absolu : nous avons trouvé dans quelques copies, pour calculer cet écart la formule :
= MAX(B2 ;C2) – MIN(B2 ;C2) ; les fonctions MAX et MIN avaient été rencontrées à l'occasion de calculs statistiques sur tableur.

Cependant une erreur rencontrée à plusieurs reprise lors de la pré-expérimentation a été l'oubli du calcul de la somme des écarts. Cette erreur est-elle à rapprocher d'une erreur fréquente en programmation dans un langage impératif qui consiste à oublier d'insérer dans le programme une instruction permettant de faire afficher ou imprimer le résultat que l'on veut calculer, voire de le faire calculer ? Cette erreur n'aurait pas résisté à un passage effectif sur machine.

3. Analyse a priori de la question B : « comment aider M. Dupond ? »

La pré-expérimentation ne nous donnera d'indications pour mener cette analyse que pour les élèves participant à l'expérimentation qui traiteront cette question en s'appuyant sur la seule simulation, puisque dans le devoir en classe, il était précisément demandé de s'appuyer sur la simulation.

Nous avons vu cependant avec l'un des extraits de réponse proposés au chapitre précédent qu'une élève avait manifesté quelque répugnance à s'appuyer sur des simulations comme le demandait l'énoncé, indiquant qu'elle aurait préféré s'appuyer sur un calcul : nous avons extrapolé de son propos qu'elle pensait à un calcul dans le modèle probabiliste, sans d'ailleurs pouvoir imaginer comment elle aurait pu « *calculer le nombre de cases minimum pour être sûr d'avoir fini avant le 25 ème tour.* »

3.1. Résolution du problème

La durée de la partie ne doit pas excéder 30 minutes. Il y a deux ou trois joueurs. Il est indiqué dans l'énoncé qu'on évalue à 1,5 minutes la durée d'un tour complet avec trois joueurs, ce qui implique qu'une partie avec trois joueurs ne doit pas compter plus de 20 tours.
Si la durée d'un tour complet est proportionnelle au nombre de joueurs, la durée d'un tour complet avec deux joueurs ne sera que d'une minute, ce qui implique qu'une partie avec deux joueurs ne doit pas compter plus de 30 tours.

Pour trois joueurs : si l'on appelle X_1, X_2 et X_3 le nombre de cases avancées par chacun des

trois joueurs pour un tour et Y_1, Y_2, Y_3 le nombre de cases avancées par chacun d'eux pour 20 tours, on calcule aisément l'espérance de X_i à partir de la loi de probabilité calculée dans la partie A. de l'énoncé : $E(X_i) = 1{,}25$, et $E(Y_i) = 20\ E(X_i) = 25$.
Le gagnant est le joueur qui va le plus loin. La question est donc de déterminer l'espérance de la variable aléatoire $Y = \max\ (Y_1\ ; Y_2\ ; Y_3)$, et ce calcul n'est pas aisé car Y peut prendre 61 valeurs, de 0 à 60. On peut affirmer cependant que $E(Y) \geq 25$.
De la même manière, pour une partie en 30 tours avec deux joueurs, le calcul de l'espérance du nombre de cases Z du trajet le plus long parcouru par le joueur qui a été le plus loin est malaisé à calculé. Cette espérance dépasse $30 \times 1{,}25 = 37{,}5$.
On peut penser que $E(Z) \geq E(Y)$ et que l'on peut se concentrer sur une partie avec trois joueurs en 20 tours.

La valeur de E(Y) est un indicateur qui ne suffit pas à prendre une décision avec une prise de risque évaluable. Le calcul de l'écart-type de Y pose les mêmes problèmes que le calcul de son espérance.
Si Y est gaussienne, il est possible de mesurer la prise de risque avec un niveau de confiance associé à un intervalle de confiance.
La simulation d'une partie avec trois joueurs peut pallier cette impossibilité de calculer directement l'espérance E(Y) et l'écart-type $\sigma(Y)$ de Y : la loi des grands nombres nous dit que la valeur moyenne et l'écart-type de la série statistique des valeurs observées pour Y lors de la répétition un grand nombre de fois de la simulation de l'expérience, convergent en probabilité vers E(Y) et $\sigma(Y)$.

Notre pré-expérimentation a fait apparaître trois difficultés d'ordres tout à fait différents :
• la première est récurrente lorsque l'on veut utiliser la convergence en probabilité de la loi des grands nombres : combien de fois faut-il répéter l'expérience pour avoir des résultats significatifs ?
Nous avons montré dans la première partie de notre thèse, chapitre 2, paragraphe 5 que la différence de convergence entre une suite numérique convergeant au sens usuel de l'analyse et une suite de fréquences convergeant en probabilité paraît être une difficulté des « programmes 2000 », qui présentent les deux notions la même année de première. Le fait que les « programmes 2010 » introduisent les intervalles de confiance avec un seuil de confiance dès la classe de seconde facilitera-t-il la distinction entre les deux types de convergence ?
Dans notre pré-expérimentation, le nombre de répétitions de la simulation proposé pour

prendre une décision, a été fort variable. Certains élèves suggéraient de faire une simulation de 25 tours pour un joueur, et de retenir le nombre de cases avancées, d'autres : plusieurs, une dizaine, cent, un très, très grand nombre de simulations.

Un passage effectif sur machine permettra peut-être aux élèves de se poser plus précisément la question du nombre de simulations à effectuer, encore que, si l'élève veut aller au delà de « plusieurs » répétitions, qui peut se faire par simple re-calcul de la feuille, la feuille de calcul devra être réaménagée (par exemple une ligne par partie de 20 tours avec trois joueurs), pour répéter par simple recopie vers le bas, la simulation un nombre conséquent de fois, afin d'utiliser ensuite les outils statistiques du tableur.

• La deuxième difficulté réside dans la prise en compte du nombre de joueurs. Lors de notre pré-expérimentation, nous avons prolongé la correction du devoir en classe par une réflexion sur l'influence du nombre de joueurs dans la solution du problème, influence qui n'allait pas de soi pour certains élèves : l'espérance de la variable aléatoire $\max(Y_1 ; Y_2 ; Y_3)$ n'est pas l'espérance commune à chacune des trois variables Y_1, Y_2 et Y_3. Nous avions pensé convaincre définitivement les élèves en leur faisant constater par eux-mêmes, sur la simulation d'une partie de 25 tours avec trois joueurs, que la valeur moyenne observée de Y ne « tournait » pas autour de 31,25 (25 × 1,25). Ce fut l'objet d'un devoir à la maison : il était demandé d'établir une feuille de calcul sur tableur pour cette simulation, et d'effectuer autant de simulations que possible.

• Et nous avons rencontré là une troisième difficulté : une faute de programmation du tableur. Certains élèves ont produit la feuille de calcul suivante (le résultat se lit dans la cellule E27, en bleu) :

	A	B	C	D	E
1	n° du lancer	Premier joueur	Deuxième joueur	Troisième joueur	
2	1	=ABS(ENT(ALEA()*4+1)-ENT(ALEA()*4+1))	=ABS(ENT(ALEA()*4+1)-ENT(ALEA()*4+1))	=ABS(ENT(ALEA()*4+1)-ENT(ALEA()*4+1))	=MAX(B2 ; C2 ; D2)
3	2	↓	↓	↓	↓
	…				
26	25				
27		=SOMME(B2 : B26)	=SOMME(C2 : C26)	=SOMME(D2 : D26)	=SOMME(E2 : E26)
28					

Au lieu de calculer la plus grande distance parcourue à l'issue des 25 tours, il est calculé dans cette feuille la somme des plus grandes distances parcourues pour chaque tour.

Une feuille de calcul correcte aurait été (le résultat se lit dans la cellule E2) :

	A	B	C	D	E
1	n° du lancer	Premier joueur	Deuxième joueur	Troisième joueur	
2	1	=ABS(ENT(ALEA()*4+1)-ENT(ALEA()*4+1))	=ABS(ENT(ALEA()*4+1)-ENT(ALEA()*4+1))	=ABS(ENT(ALEA()*4+1)-ENT(ALEA()*4+1))	=MAX(B27 ; C27 ; D27)
3	2				
	…				
26	25				
27		=SOMME(B2 : B26)	=SOMME(C2 : C26)	=SOMME(D2 : D26)	
28					

La programmation du calcul dans un langage impératif permettrait peut-être plus facilement de faire une *imitation au plus près* des actions en jeu ici, la séquence des instructions faisant peut-être mieux apparaître la chronologie des différentes actions que ne le permet le tableur.

3.2. Analyse a priori des difficultés de la résolution du problème : « comment conseiller ? »

Nous présenterons dans la quatrième partie de notre thèse une solution complète au problème posé aux élèves dans notre expérimentation, avec une interprétation dans le cadre de la théorie des probabilités des résultats obtenus par cette solution ; nous présenterons également alors l'exploitation qu'il est peut-être possible de faire de ce problème ou de problèmes analogues dans une progression du cours de probabilité.

Nous voulons revenir sur deux difficultés d'ordre théoriques qui nous apparaissent a priori.

Le calcul de l'espérance de Y

Nous avons écrit dans le paragraphe précédent, sans justifier ce résultat, que

$E(Y_i) = 20 \times E(X) = 25$.

En jouant sur les mots, il paraît de « bon sens » que d'affirmer que l'espérance de gain d'un joueur sur 20 tours est égale à 20 fois son espérance sur un tour.

Ce résultat ne peut en fait se justifier qu'en reprenant rigoureusement la définition de la variable aléatoire Y, qui est une variable aléatoire sur l'univers produit Ω^{20} de l'espace probabilisé produit, qui modélise la répétition 20 fois de l'expérience « on fait un tour avec trois joueurs ».

Pour un tour, l'espace probabilisé est :

$(\Omega, \mathcal{B}, p)$ où $\Omega = \{1 ; 2 ; 3 ; 4\}$, $\mathcal{B}$ est l'ensemble des parties de Ω, et p la loi équirépartie.

On introduit pour le premier joueur les 20 variables aléatoires $X_{1,i}$, où $X_{1,i}$ est le nombre de cases avancée lors du i-ème tour par ce premier joueur, variables aléatoires définies sur Ω^{20}. Pour tout entier i entre 1 et 20, $E(X_{1,i}) = 25$. La variable aléatoire Y_1 égale au nombre de cases avancées par ce premier joueur à l'issue de 20 tours est définie également sur Ω^{20}, et est en fait la somme sur l'indice i : $Y_1 = \sum X_{1,i}$; son espérance est, par linéarité de l'espérance : $E(Y_1) = \sum E(X_{1,i}) = \sum 1{,}25 = 25$.

Quels indicateurs statistique choisir ? Quel risque accepter ?

Dans le cas d'une étude statistique de la série des *n* valeurs observées pour $Y = \max (Y_1 ; Y_2 ; Y_3)$ lors d'une répétition *n* fois de la simulation, se pose la question du choix des indicateurs à calculer : dans notre pré-expérimentation, certains élèves ont retenu la plus petite valeur de la série, d'autres la valeur de fréquence maximale, d'autres la valeur moyenne. Comme nous l'avons évoqué à propos de l'espérance de Y, la valeur moyenne de la série statistique des Y observés ne suffit pas ; le calcul de l'écart-type de la série statistique ne pose pas de problème, mais il reste à savoir comment l'utiliser, ce qui n'est pas un résultat des « programmes 2000 » des classes scientifiques : les élèves doivent savoir que pour une même valeur moyenne *m*, un plus ou moins grand écart-type indiquera une plus ou moins grande dispersion des séries statistiques autour de cette valeur moyenne commune ; mais ici, une seule série statistique est en jeu, la série des valeurs de Y observées, donc il n'y a pas de comparaison possible avec une autre série.

En fait il s'agit d'estimer le risque à associer à la décision qui sera prise. Cette problématique ne figure pas dans les « programmes 2000 » de première S. Par contre, elle peut être rapprochée du point « adéquation de la loi équirépartie aux données expérimentales » du programme de terminale S, et de la solution qui en est proposée dans ces programmes (voir chapitre 22, paragraphe 1.2.), solution fondée justement sur les simulations.

Nous n'attendons pas a priori dans notre expérimentation de proposition de cette sorte, d'autant que le test d'adéquation que nous venons d'évoquer se traite en général en fin d'année scolaire, et que nous pensons faire notre expérimentation au deuxième trimestre, c'est pourquoi nous repoussons à la quatrième partie de notre thèse la description détaillée d'une solution complète avec estimation du risque associé à la prise de décision dans le problème de M. Dupond.

Par contre nous avons trouvé dans notre pré-expérimentation des élèves (voir par exemple l'un des trois extraits du chapitre précédent) qui souhaitaient se donner une certaine marge de manœuvre par rapport à l'indicateur calculé, manifestant par là qu'ils ont vu que l'indicateur qu'ils proposent de calculer ne donne pas de certitude, et qu'il convient de l'adapter pour rendre « plus sûre » la décision.

4. Les variables didactiques potentielles de notre expérimentation

Nous considérons qu'une variable de l'expérimentation est une variable didactique lorsque nous pouvons changer sa valeur <u>et</u> lorsqu'elle est susceptible d'avoir un effet sur le résultat du travail de l'élève, ou du groupe d'élèves.

<u>4.1. Avec ou sans ordinateur ?</u>

Nous pensons que la possibilité offerte aux élèves d'utiliser un ordinateur est potentiellement une variable didactique de notre expérimentation. Cette hypothèse est fondée en particulier sur l'impression que devant l'ordinateur, les élèves qui avaient participé à la pré-expérimentation <u>sans ordinateur</u>, auraient peut-être formulé différemment leurs réponses à la dernière question du devoir (combien de cases choisir ?) en réalisant <u>effectivement</u> la simulation, ne serait-ce qu'en proposant une ou des réponses numériques. Nous avons donc organisé le dispositif expérimental dans un cadre idéal où l'élève est à tout moment libre d'utiliser l'ordinateur.

<u>4.2. L'ordre des questions</u>

L'ordre des questions est également une variable didactique potentielle.
Rappelons que son effet peut être double :

- ce peut être un effet sur le choix de la technique de simulation choisie par l'élève : un élève qui a d'abord déterminé la loi de probabilité de la variable X égale au nombre de cases avancée par un joueur sur un tour, va-t-il simuler cette loi de probabilité, ou faire une *imitation au plus près* ?
- Ce peut être un effet sur la réponse de l'élève à la question B. : va-t-il choisir d'utiliser la modélisation probabiliste plutôt qu'une simulation selon qu'il aura dans la question qui précède immédiatement déterminé la loi de probabilité ou la simulation demandée, ou la question qui précède immédiatement est-elle sans effet ?

5. En conclusion : les réponses attendues

Voici les grandes lignes des réponses que nous attendons pour chacune des questions.

5.1. Algorithme pour la simulation

Nous attendons quatre types d'algorithmes fondés sur :

(1) : deux tirages d'un entier entre 1 et 4 selon une loi équirépartie pour les élèves qui font une imitation « au plus près » en imitant *les expériences élémentaires* composant l'expérience globale ;

(2) : le tirage d'un entier entre 0 et 3 selon une loi équirépartie pour les élèves qui font une imitation des *événements intéressants* : « on avance de 0 » , « on avance de 1 », « on avance de 2 », « on avance de 3 », avec le biais d'équiprobabilité ;

(3) : le tirage d'un entier aléatoire entre 1 et 8 ou 1 et 16 (loi de probabilité correcte, avec ou sans simplification des fractions) selon une loi équirépartie pour les élèves qui font une simulation d'un tirage d'urnes après avoir déterminé la loi de probabilité de l'expérience aléatoire ;

(4) : le tirage d'une entier entre 1 et 10 selon une loi équirépartie (même démarche qu'en **(3)**, mais la loi de probabilité est erronée car les issues ont été représentées par des paires).

Les algorithmes **(3)** et **(4)** seront choisis par les élèves qui simulent la loi de probabilité en simulant le tirage d'un boule d'une urne de composition bien choisie, expérience équivalente à l'expérience initiale, puisqu'ayant même loi de probabilité. C'est la technique que préconisent les programmes et les documents d'accompagnement.

5.2. Programmation de la simulation

Nous attendons quatre types de feuilles de calcul, le cadre imposé sur l'énoncé orientant vers l'organisation : une ligne du tableur pour un tour.

(type 1) : une cellule pour chaque dé, l'imitation s'étalant ensuite sur plusieurs cellules à droite :

ALEA .ENTRE.BORNES(1 ; 4)	ALEA .ENTRE.BORNES(1 ; 4)	………	………..

Les flèches indiquent une recopie des cellules vers le bas.

(type 2) : une formule globale dans une cellule telle que :

=ABS(ALEA .ENTRE.BORNES(1 ; 4) - ALEA .ENTRE.BORNES(1 ; 4)),

pour ces deux premiers types, la ligne, ou la cellule, simulant un lancer des deux dés et le calcul de l'écart absolu des points amenés est recopiée vers le bas sur 20 lignes,

le total des cases avancées pourra être calculé en ligne 22 ou 23, ou ligne 1 dans la colonne la plus à droite, en utilisant la fonction SOMME.

(type 3) : simulation d'un tirage d'une boule d'une urne contenant des boules de caractéristiques diverses en lien avec la loi de probabilité déterminée.

Par exemple, tirage d'une urne contenant 8 boules numérotées de 1 à 8 pour une loi de probabilité correcte ramenée au dénominateur 8 :

x_i =	0	1	2	3	total
$p(X = x_i)$	2/8	3/8	2/8	1/8	1

en A1 (ligne 1, colonne A) : =Alea.entre.bornes(1 ; 8)

en B1 (ligne 1, colonne B) : =SI(A1<3 ; 0 ;SI(A1<6 ;1 ;SI(A1<8 ;2 ;3)))

c'est à dire : *si A1 < 3 alors 0 sinon si A1 < 6 alors 2 sinon 3.*

Les deux cellules de cette première ligne sont alors recopiées sur 20 lignes, et comme pour les types (1) et (2), le total des cases avancées est calculé en ligne 22 ou 23, ou ligne 1 dans la colonne la plus à droite.

(type 4) en A1 : =Alea.entre.bornes (0 ; 3) pour une imitation des *événements intéressants* 0, 1, 2 et 3 selon une loi (erronée) équirépartie, la formule étant alors recopiée sur 20 lignes vers le bas, et le total des cases avancées calculé comme dans les autres feuilles.

5.3. Modélisation

Nous attendons trois lois de probabilité :

• Loi équirépartie sur {0 ; 1 ; 2 ; 3}, correspondant au biais d'équiprobabilité.

• Loi correcte, obtenue avec un arbre, ou un tableau, ou par cas (X=0, X=1, etc.), comme nous l'avons décrit plus haut.

• Loi fausse, mais les issues 0, 1, 2 et 3 ne sont pas équiréparties ; il s'agit du cas où l'élève ne « distingue » pas bien les deux dés en les représentant par des paires.

3.4. Réponses attendues à la question B.

(1) On fait une simulation de 20 tours, on prend le résultat observé.

(2) On fait plusieurs simulations de 20 tours, on prend la moyenne des résultats observés.

(3) On répète un très grand nombre de fois la simulation de 20 tours, on prend la moyenne des résultats observés.

(4) On répète un très grand nombre de fois la simulation de 20 tours, on prend la plus petite valeur observée.

(5) On répète un très grand nombre de fois la simulation de 20 tours, on prend la valeur observée qui a la plus grande fréquence.

(6) On calcule l'espérance de la variable aléatoire égale au nombre de cases avancées pour un lancer des deux dés, que l'on multiplie par 20.

Peut-être des élèves envisageront-ils de retirer quelques cases aux valeurs présentées ci-dessus, « par sécurité ».

Peut-être des élèves essayeront-ils de tenir compte du nombre de joueurs.

Nous n'attendons pas de réponse formulée en termes de risque pour Monsieur Dupond.

CHAPITRE 15

Le dispositif expérimental

157 élèves ont effectué le travail pendant une séance de travaux dirigés, donc en demi classes. Ces élèves proviennent de deux classes de terminale S et 3 classes de première S de deux lycées. Trois enseignantes ont été concernées par cette expérimentation.

Nous avons changé le dispositif prévu à l'origine : nous avions prévu d'offrir la possibilité à chaque élève d'utiliser à sa convenance un ordinateur, dans une situation idéale avons-nous dit au chapitre précédent, où chaque élève disposerait d'un ordinateur.
Deux classes ont participé à l'expérimentation en 2008-2009 : une classe de première S, à qui a effectivement été offerte la possibilité d'accéder aux ordinateurs, et une classe de terminale S dont les élèves n'ont pas eu cette possibilité, en raison de la « disparition » d'un grand nombre de souris, disparition constatée juste avant que ne démarre l'expérimentation.

Après avoir comparé les réponses des deux classes, nous avons alors décidé, pour les classes qui ont composé en 2009/2010, de repousser à une autre séance ou une autre forme d'organisation – à l'initiative de l'enseignante, et en dehors de notre expérimentation – le travail sur ordinateur, pour trois raisons :

• Les salles informatiques des deux lycées concernés ne permettaient pas en réalité d'effectuer le travail dans les conditions souhaitées : les ordinateurs y sont disposés en pourtour de salle (les élèves se déplaçant pour les utiliser) et sont en nombre insuffisant ; l'impossibilité d'affecter un ordinateur à chaque élève, rendait problématique le caractère individuel du travail ; l'enseignante du groupe ayant utilisé des ordinateurs en 2008-2009 avait de plus noté un « effet de groupe » : lorsqu'un élève se lève pour aller aux ordinateurs, les autres ont tendance à l'imiter.

• Les élèves devaient tous travailler dans les mêmes conditions : nous avons donc décidé de ne pas intégrer à l'analyse d'ensemble les observations liées aux travaux des élèves de la classe de première S qui avait eu la possibilité d'utiliser des ordinateurs, et par contre de conserver les travaux de la classe de terminale S qui a composé en 2008/2009.

• Mais surtout nous avions retrouvé, avec le groupe ayant travaillé avec ordinateur, ce que les enseignants de l'option informatique des années 1980-1990 avaient pu observer : un élève placé immédiatement devant l'ordinateur a tendance a vouloir programmer tout de suite, au détriment de la réflexion sur le problème qu'il a à résoudre, en se concentrant souvent sur des questions de mise au point sans importance (du moins au départ). C'est, pensons-nous, ce qui explique que les élèves ayant travaillé sur ordinateur n'ont pas expliqué leur démarche, mais ont, pour la plupart, seulement donné une réponse numérique non commentée. Or nous souhaitions avant tout étudier des réponses argumentées, d'autant que la résolution complète du problème, qui passe certes probablement par une mise en œuvre de simulations, nous paraissait dépasser largement ce qu'on peut attendre d'un élève seul en temps très limité.

Cette décision nous a amenée à écarter les résultats de toute la classe de première S qui avait composé en 2008-2009 : même si aucun élève du deuxième groupe n'a jugé utile d'utiliser la faculté qui était offerte d'utiliser les ordinateurs, cette faculté n'était pas offerte aux autres classes. L'effet de groupe décrit par l'enseignante qui a vu tout un groupe se rendre aux ordinateurs parce qu'un élève avait mis en route un ordinateur, a pu jouer dans l'autre sens : « on ne va pas y aller puisque personne n'a l'air de juger utile d'y aller ».

Nous analyserons donc les réponses d'un ensemble de 134 élèves répartis sur deux classes de première S et deux classes de terminale S d'un même lycée.

La passation s'est effectuée par demi classe.
Le choix de mener notre expérimentation par demi classe était à l'origine une nécessité pour permettre un passage sur machine. Nous avons conservé l'organisation par demi classe pour maintenir la même organisation dans toute les classes (la classe de terminale S ayant composé en 2008-2009 a composé par demi classe) ; cette organisation de plus facilitait la répartition des deux modalités de l'énoncé : lorsque dans une classe un groupe composait avec l'énoncé 1 (modélisation avant la simulation), l'autre groupe composait avec l'énoncé 2 (simulation avant modélisation).

Nous avons conservé dans chaque classe la composition habituelle des groupes de travaux dirigés de mathématiques. Ces groupes sont constitués soit selon l'ordre alphabétique, soit selon l'ordre alphabétique et des considérations de compatibilité d'emploi du temps lorsque les travaux dirigés de mathématiques ont lieu « en parallèle » avec un enseignement dans une

autre discipline, en l'espèce la deuxième langue vivante.
Nous faisons l'hypothèse que dans une classe donnée, chaque groupe est représentatif de la classe entière, ce qui paraît une hypothèse raisonnable vus les critères de constitution des groupes. Cette hypothèse assure que dans chaque classe, les deux modalités des énoncés seront distribuées à deux groupes comparables.

Il est difficile de comparer les deux classes de terminale en dehors du constat qu'elles ont obtenu un taux comparable de réussite au baccalauréat.
Les deux classes de première S ont fait un devoir commun de mathématiques en fin d'année scolaire : l'une des classes a présenté des indicateurs statistiques concernant les notes obtenus par chaque élève plus « favorables » que l'autre.

Sous l'hypothèse que dans chaque classe, les deux groupes de travaux-dirigés sont représentatifs de la classe, cette hétérogénéité des deux classes de première S n'a pas d'incidence sur la composition des groupes d'élèves ayant composé avec l'un ou l'autre des énoncés : ces groupes sont comparables.

Chaque groupe était constitué d'au plus 18 élèves, disposés dans une salle prévue pour accueillir 35 élèves ; cela nous a permis de séparer les élèves pour assurer un travail personnel.

Le travail a été présenté aux élèves comme un travail individuel de réflexion, corrigé par le professeur, mais non noté. Dans chaque groupe de travaux-dirigés, l'enseignante a expliqué comment on lit le numéro sorti sur un dé tétraédrique, et un dé tétraédrique a circulé dans chaque groupe. Cette présentation du travail et du dé tétraédrique a duré moins de 5 minutes.

Les élèves ont donc disposé d'environ 50 minutes, la durée habituelle d'une séquence de cours étant de 55 minutes.

Chaque élève a reçu une liasse de trois feuilles : la première page contenait une partie commune aux deux énoncés et une partie différenciée correspondant aux deux modalités (<u>énoncé 1</u> : la simulation après la loi de probabilité et <u>énoncé 2</u> : la simulation avant la loi de probabilité) ; la deuxième page comportait les cadres dans lesquels les élèves devaient répondre aux question, cette page étant bien sûr également différenciée selon la modalité du travail ; enfin la troisième page, commune aux deux énoncés, contenait un memento des principales commandes du tableur « open office ».

Les données de notre expérimentation se trouvent donc sur la page 2 des énoncés, sur laquelle les élèves ont composé leurs réponses.

Dans les deux classes de première S, la loi de probabilité de la variable aléatoire égale à la somme des points amenés lorsqu'on lance deux dés cubiques équilibrés avait été traitée en exercice pour introduire la notion de loi de probabilité d'une variable aléatoire.
Dans ces deux classes, la solution proposée pour cet exercice s'appuyait sur la représentation des issues à l'aide d'un tableau à deux entrées ; cette solution a été décrite dans le chapitre précédent.
Le travail de notre expérimentation a été proposé à ces deux classes de première S à l'issue du cours de probabilité.

Dans les deux classes de terminale S, l'exercice de détermination de la loi de probabilité de la somme des points amenés lorsqu'on lance deux dés équilibrés a été proposé aux élèves lors d'une séance de révision du cours de probabilités de première S (la simulation d'un dé équilibré à six faces a été revue lors de cette séance).
Dans ces deux classes, dont nous étions l'enseignante, cette séance de révision du cours de première S, a précédé le chapitre de dénombrement du programme de terminale S. Les exercices de dénombrement proposés aux élèves ont souvent été des exercices de calcul de probabilités, exercices en réalité prétextes à mener des dénombrements : ainsi par exemple l'exercice « on tire au hasard une main de cinq cartes d'un jeu de 32 cartes, quelle est la probabilité que cette main contienne une tierce mais pas un cinquante [66] ? ».
Nous avons inséré comme à notre habitude plusieurs chapitres entre ces révisions du programme de probabilités de première S précédant le chapitre de dénombrement, et les chapitres de probabilités du programme de terminale S.
Notre expérimentation dans ces deux classes a précédé immédiatement ces chapitres de probabilité du programme de terminale. Notre expérimentation s'est donc déroulé à une certaine distance des révisions de première S, et les élèves de terminale avaient le même « bagage théorique » en probabilités que les élèves de première pour aborder le problème de

[66] « tierce » et « cinquante » sont des annonces au jeu de la belote.

notre expérimentation.

Précisons les effectifs, et la ventilation des énoncés qui, rappelons-le, ne différaient dans leur présentation, que par l'ordre des deux questions posées dans la partie A :

niveau / sujet	première	terminale	total
1 (modélisation avant simulation)	44	33	77
2 (simulation avant modélisation)	46	34	80
total	90	67	157

Nous avons repéré chaque élève par un numéro. Nous désignerons dans la suite chaque élève par ce numéro.
Nous avons transcrit, en annexe du chapitre 18 , l'intégralité des réponses à la question B. , y compris les réponses de la classe écartée de l'étude qui va suivre (ce sont les réponses affectées des numéros 45 à 56).

En conclusion, le plan expérimental de notre expérimentation comporte deux variables indépendantes : le niveau scolaire (première ou terminale) et l'énoncé (énoncé 1 : la modélisation précède la simulation ou énoncé 2 : la simulation précède la modélisation).

CHAPITRE 16

Les réponses des élèves : classifications

1. Une première classification des travaux des élèves

Cette classification concerne essentiellement les réponses à la première partie du sujet (simulation de 20 tours pour un joueur et détermination de la loi de probabilité de l'écart lors du lancer des deux dés). Nous avons également regardé s'il y avait une réponse pertinente à la question B, et dans l'affirmative, si elle s'appuyait sur les simulations ou sur le modèle probabiliste, ou les deux, ou sur d'autres considérations.

1.1. Classification des réponses pour la simulation demandée

Certaines réponses ont été décrites dans l'analyse a priori. Dans un souci de clarté, nous les reprenons ici.

Les regroupements de réponses

Nous avons pu classer la réponse de chaque élève dans l'un des groupes suivants :

• **loi de probabilité** : la simulation est correcte. La feuille de calcul simule la loi de probabilité associée à la variable X, par le détour d'un tirage d'urne (cf. l'analyse de cette méthode dans les chapitres précédents).

Un seul élève est concerné.

• **imitation** : la simulation est correcte. Il s'agit d'une *imitation au plus près* du jeu, avec deux dés tétraédriques équilibrés. Le nombre de cases parcourues à l'issue des 20 tours est calculé.

Nous intégrons dans ce groupe deux élèves qui ont travaillé avec des dés cubiques équilibrés à 6 faces.

Nous avons également classé dans ce groupe un élève qui a changé la règle du jeu : on lance les deux dés ; s'il sort un double, le joueur n'avance pas, sinon, le joueur avance d'autant de cases que la somme des points amenés sur les deux dés. La simulation et la loi de probabilité que cet élève propose sont correctes pour cette règle du jeu.

Au total, nous avons classé 81 élèves dans ce groupe.

• « **Dé5** » : simulation de deux dés à cinq faces par Alea.entre.bornes (0 ; 4) ; nous regrouperons dans les tableaux statistiques cette réponse avec les réponses classées « **imitation** », car en dehors de cette erreur, la feuille de calcul est correcte. *(un élève)*.

• **imitation incomplète** : L'imitation du jeu est correcte (cf la rubrique « **imitation** »), mais l'élève ne fait pas calculer le nombre total de cases parcourues après 20 tours.

Il en va ainsi pour 15 élèves.

• **imitation des lancers** : la feuille de calcul présente uniquement une simulation du lancer de deux dés tétraédriques. L'écart des points amenés n'est pas calculé. Nous considérerons que ces élèves ont correctement exploité les hypothèses de modélisation sur les dés.

5 élèves sont dans ce cas.

• **loi équirépartie sur [0 : 3]** : la feuille de calcul proposée correspond à la simulation d'une loi équirépartie sur {0 ; 1 ; 2 ; 3}. La variable aléatoire « nombre de cases avancées peut prendre les valeurs entières de 0 à 3, mais ces valeurs n'ont pas la même probabilité, cette imitation des écarts est donc fausse.

Nous analysons cette erreur, rencontrée également lors de la pré-modélisation, comme une erreur dans l'étape de pré-modélisation.

Nous avons intégré à ce groupe un élève qui a travaillé avec deux dés à 6 faces, et qui a produit comme réponse une loi équirépartie sur {0 ; 1 ; … ; 5}.

12 élèves ont commis cette erreur.

• **Autre** :

Nous avons classé dans ce groupe des élèves pour lesquels :

- le jeu n'est pas compris,
- ou la feuille de calcul nous paraît sans rapport avec le problème posé,
- ou le fonctionnement d'un tableur n'est pas compris :

au total les réponses de 5 élèves ont été classées dans ce groupe.

• **néant** : l'élève n'a pas répondu à la question.

14 élèves sont dans ce cas.

Le calcul de l'écart

Le calcul de l'écart absolu entre les numéros sortis sur les dés a été fait selon diverses procédures. Nous pensons utile de revenir sur ces procédures car il semble que ce calcul a posé quelques problèmes à certains élèves.

Dans notre expérimentation, la fonction Abs était rappelée dans le mémento fourni avec l'énoncé ; certains élèves n'ont peut-être pas pensé à chercher dans ce mémento la fonction valeur absolue, reconstruisant cette fonction de diverses manières.

Nous avons dans l'expérimentation retrouvée deux procédures décrites dans la pré-expérimentation (dans le cas où la simulation de chacun des deux dés se fait dans les cellules B2 et C2), et une nouvelle procédure :

• ABS(B2-C2) ;

• MAX(B2 ; C2) – MIN(B2 ; C2) , rencontrée deux fois ;

• SI(C2 > B2 ; C2 – B2 ; B2 – C2), rencontrée 10 fois.

Ces deux dernières procédures peuvent être regardées comme une imitation très proche de ce que fait le joueur lorsqu'il veut calculer l'écart : il compare les deux numéros sortis pour repérer le plus grand, etc.

Mais peut-être ces deux procédures manifestent-elles en réalité une difficulté pour certains élèves à « traduire » l'énoncé en termes mathématiques avec une valeur absolue :

« *il* [le joueur] *avance sur le plateau d'un nombre de cases égal à l'écart entre les points amenés par les deux dés (dans le cas d'un double, le joueur n'avance pas* ».

La copie n° 70 semble manifester cette difficulté :

L'élève propose deux simulations :

• elle propose dans la cellule A1 la formule =ALEA.ENTRE.BORNES(0 ;3) avec le commentaire « *soustraction des valeurs obtenues par les deux dés* », qui est donc une imitation des *événements intéressants* avec des hypothèses de modélisation fausses : les écarts 0, 1, 2 et 3 ne sont pas équiprobables.

• Mais elle écrit en dessous du tableau :

> *« On peut utiliser ALEA.ENTRE.BORNES(0 ; 3), cette formule nous donne directement la soustraction obtenue par les 2 dés. Sinon il aurait fallu 3 formules de plus : ALEA.ENTRE.BORNES(1 ; 4) deux fois, et ~~faire~~ calculer la soustraction des deux chiffres obtenues en prenant d'abord le plus grand ».*

Cette autre proposition semble correspondre à une *imitation au plus près* de l'expérience, avec des hypothèses de modélisations correctes.

On peut penser , à lire la deuxième proposition, que la première proposition est venue après, parce que l'élève était en difficulté pour faire calculer l'écart absolu entre les points amenés, n'ayant peut-être jamais utilisé la fonction ABS du tableur, et ne l'ayant pas vu sur le memento. Un indice de cette difficulté pour calculer l'écart est donné par son explication « *calculer la soustraction des deux chiffres obtenues en prenant d'abord le plus grand* » : comment prendre d'abord le plus grand était peut-être justement la question que cet élève ne savait pas résoudre.

L'élève présente sa solution comme une simplification de la feuille de calcul (« *sinon il aurait fallu trois formules* », sans réaliser que la deuxième feuille ne met pas en jeu les mêmes hypothèses de modélisation.
Nous avons classé cette réponse à la simulation dans le groupe « **loi équirépartie sur [0 : 3]** ».

Deux réponses difficiles à classer

• Nous classons également dans ce groupe la copie de l'élève n° 78 dont nous reproduisons ci-dessous la copie :

A. 1. Feuille de calcul du tableur

	A	B	C	D
1	ALEA.ENTRE.BORNES (0;3)	ALEA.ENTRE.BORNES (0;3)		
2				
20				
21				
22				

A. 2.

X =	0	1	2	3
probabilité correspondante	1/16	1/16	1/16	1/16

Justifications : $P(0) = P(1) = P(2) = P(3) = \frac{1}{16}$ car on a 2 dés équilibrés, donc la loi est équirépartie et on a 4 issues possibles par dé : 0 ; 1 ; 2 et 3. Donc ce qui fait 4x4 = 16 issues.
Donc la probabilité est la même pour chacune des issues et vaut $\frac{1}{16}$

B. Comment conseiller Monsieur Dupond ?

La formule de la cellule A1 semble indiquer que cette élève considère que la loi de X est équirépartie sur l'ensemble {0 ; 1 ; 2 ; 3} (rappelons qu'un dé tétraédrique circulait dans la classe, et que les élèves ont pu constater que ses faces sont numérotées de 1 à 4), mais comment interpréter la formule en B1 : comme la volonté de faire intervenir un second joueur, ou les deux cellules incarnent-elles chacun des deux dés ? Nous n'avons pu déchiffrer ce qui a été recouvert par du blanc dans la cellule C1.

Le rapprochement de la réponse proposée pour la simulation avec la réponse proposée pour la loi de probabilité nous a amenée à considérer que les deux réponses relevaient du biais d'équiprobabilité. La réponse à la question A2 paraît illustrer le passage mal conduit des hypothèses de modélisation sur chacun des deux dés, à la conclusion que les *issues physiques*, (correctement dénombrées par cette élève) sont équiprobables : l'élève a fait porter cette conclusion sur les *événements qui l'intéressent*.

• L'élève n° 150 présente des réponses assez comparables (notons que ces deux élèves ne proviennent pas de la même classe) :

A. 1.

X =	0	1	2	3
probabilité correspondante	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Justifications :

Je sais que le dé tetraedrique est equilibré
Dans ce cas la somme des probabilité est egale à 1.
Il y a donc 4 evenement A,B,C,D possible : Paire 0, 1, 2, 3.
La regle stipule que X est l'ecart des points amené par les 2 dés a quatre faces.
$U = 4 \times 4 = 16$
"A" = (1,1) ; (2,2) ; (3,3) ; (4,4) donc $p(A) = \frac{\text{card } A}{\text{card } U} = \frac{4}{16} = \frac{1}{4}$
De meme pour les evenements B, C et D

A. 2. Feuille de calcul du tableur

	A	B	C	D
1	ALEA.ENTRE.BORNES(0;3)	ALEA.ENTRE.BORNES(0;3)		
2	ALEA.ENTRE.BORNES(0;3)	ALEA.ENTRE.BORNES(0;3)		
20	ALEA.ENTRE.BORNES(0;3)	ALEA.ENTRE.BORNES(0;3)		
21	SOMME(A1:A20)	SOMME(B1:B20)		
22				

B. Comment conseiller Monsieur Dupond ?

Monsieur Dupond peut utiliser la formule MAX dans son tableur pour savoir le nombre de point maximum qu'un joueur peut obtenir. Le nombre de cases de son jeux sera donc inferieur ou egale à ce nombre maximum.

L'élève semble avoir bien vu que lorsqu'on lance deux dés tétraédriques, il y a 16 issues équiprobables, et qu'il y a quatre événements intéressants correspondant aux quatre valeurs possibles pour l'écart ; il a correctement mené le calcul de la probabilité d'avoir un écart égal à zéro, en faisant la liste des cas favorables.

On peut donc considérer que ses hypothèses de modélisation dans l'étape de pré-modélisation semblent correctes.

Ensuite, sans détailler plus les événements, B, C et D, il les déclare de même probabilité, comme si ces événements étaient devenus les issues.

La simulation, qui suit la modélisation, pourrait être interprétée comme la simulation de la loi de probabilité que l'élève vient de déterminer : les valeurs 0, 1, 2 et 3 selon la loi uniforme.

Mais alors pourquoi la colonne B, sinon pour simuler un deuxième dé ? Certains élèves (peu) ont simulé le dé tétraédrique par Alea.entre.bornes(0 ; 3), ce qui ne pose pas de problème sur la qualité de la simulation puisque après est calculé l'écart, mais une telle interprétation du choix d'Alea.entre.bornes(0 ; 3) ne va pas de soi ici.

Cependant l'élève a aussi calculé la somme de la colonne B. Est-ce qu'en fait la colonne B ne serait pas plutôt là pour prendre en compte un deuxième joueur ?

A moins que la colonne A ne représente une première simulation (erronée) de 20 tours, et la colonne B une seconde simulation.

Cette troisième interprétation éclairerait la réponse de cet élève pour la question **B.**, puisqu'il propose d'utiliser la commande MAX (Max de A21 et de B21 ?).

Nous avons décidé de classer également ces deux réponses dans les groupes correspondant au biais de l'équiprobabilité sur {0 ; 1 ; 2 ; 3}.

Nous veillerons à tenir compte de la difficulté à classer ces trois copies dans le groupe « **loi équirépartie sur [0 : 3]** » dans l'analyse que nous ferons de ce groupe.

1.2. Classification des réponses pour la loi de probabilité

Nous avons rencontré diverses réponses, correctes ou non, pour la loi de probabilité.

Nous avons rencontré quelques réponses nouvelles par rapport à la pré-expérimentation. Nous avons également pu classer chaque réponse dans l'un des groupes décrits ci-dessous. Là encore, dans un souci de clarté, nous rappelons les réponses attendues dans l'analyse a priori.

• **correcte** : la modélisation est correcte. Le tableau de probabilité proposé est :

x_i	0	1	2	3
$p(X=x_i)$	$\frac{4}{16}$ ou $\frac{1}{4}$	$\frac{6}{16}$ ou $\frac{3}{8}$	$\frac{4}{16}$ ou $\frac{1}{4}$	$\frac{2}{16}$ ou $\frac{1}{8}$

Nous avons intégré à ce groupe les réponses de trois élèves qui ont travaillé avec un dé à 6 faces et ont produit une loi de probabilité correcte pour ces dés.

Les réponses de 108 élèves rentrent dans ce groupe.

• **paires** : l'élève n'a pas distingué l'issue (1 ; 4) et l'issue (4 ; 1). La loi de probabilité proposée est alors

x_i	0	1	2	3
$p(X=x_i)$	$\frac{4}{10}$	$\frac{3}{10}$	$\frac{2}{10}$	$\frac{1}{10}$

14 réponses ont été classées dans ce groupe.

• **8 doubles** : la sortie d'un 1 et d'un 4 est bien comptée deux fois (1 et 4, 4 et 1) mais le double 4 aussi ; il y a alors 8 issues correspondant à des doubles, et donc au total 20 issues, parmi lesquelles 8 doubles.

2 élèves ont fait cette erreur.

• **nombre de cas favorables** : l'élève, dans le tableau de la loi de probabilité, donne le nombre de cas favorables à X = 0, X = 1, etc., et non les probabilités qu'on obtient en divisant ces nombres de cas favorables par le nombre de cas possibles : 16.

3 élèves sont dans ce cas.

• **loi équirépartie sur [0 : 3]** : l'élève propose la loi équirépartie sur l'ensemble des entiers {0 ; 1 ; 2 ; 3}

4 élèves sont dans ce cas.

• **Autre** : Il s'agit de copies pour lesquelles le jeu n'est pas compris, et dont les réponses ne peuvent être interprétées en termes de compréhension des hypothèses de modélisation, à la différence des trois copies qui ont travaillé avec des dés cubiques.

2 élèves sont dans ce cas.

• **Néant** : pas de réponse

1 élève est concerné.

1.4. Classification des réponses correspondant à la question **B.**

Le conseil s'appuie-t-il sur la simulation de 20 lancers ou sur la modélisation probabiliste ou sur une autre approche ?

Nous avons là encore classé les réponses dans plusieurs groupes :

• **néant** : pas de réponse

C'est le cas pour 23 élèves.

• **Simulation** : le conseil s'appuie sur la simulation exclusivement.

29 réponses sont dans ce cas.

• **Modélisation** : le conseil s'appuie exclusivement sur la modélisation probabiliste, c'est à dire la loi de probabilité pour un lancer des deux dés.

50 réponses sont classées dans ce groupe.

• **Modélisation et simulation** : le conseil s'appuie à la fois sur la modélisation et la simulation.

Cela correspond à 5 réponses.

• **procédure incertaine** : La réponse est relativement pertinente, et repose probablement sur la modélisation, mais l'élève n'a pas laissé assez d'indices pour que nous en soyons certaine.

C'est le cas pour un élève.

• **Autre :** nous classons dans ce groupe des réponses variées :

- des réponses qui semblent sans rapport avec la question posée qui est « comment aider M. Dupond ? », y compris lorsqu'est calculée l'espérance du nombre X de cases avancées par un joueur pour un tour si ce seul calcul constitue la réponse proposée ;
- une réponse où est affirmée l'impossibilité de répondre à la question car cela dépend du nombre de joueurs ;
- des réponses consistant dans le simple constat qu'il faut mettre au plus 60 cases.
- quelques réponses qui paraissent incohérentes.

Ce groupe comporte 26 réponses.

Nous reportons au paragraphe qui suit une présentation détaillée des conseils proposés pour choisir le nombre de cases, ainsi que leur analyse.

2. Une deuxième classification pour la question B. (comment conseiller ?)

Cette classification ne concerne que les réponses que nous n'avons pas classées dans les groupes « **autre** » ou « **néant** ».

Nous avons défini un certain nombre de caractères que possède ou ne possède pas la réponse d'un élève et qui nous ont paru significatifs pour le deuxième axe d'observation que nous nous étions assigné : comment les élèves vont-il essayer de « conseiller Monsieur Dupond », et, s'ils essayent d'investir des outils tels que la simulation ou la modélisation probabiliste, quelles procédures imagineront-ils ?
Les caractères retenus sont qualitatifs, à l'exception de deux d'entre eux.
Ces caractères ne définissent pas des groupes disjoints

4.1. Caractères liés à la modélisation probabiliste

caractère	description de ce caractère
Les élèves qui se sont appuyés sur la loi de probabilité de la variable X égale au nombre de cases avancées par un joueur, sur un tour , ont tenté de calculer de combien de cases le joueur peut *espérer* avancer sur 20 tours.	
calcul de l'espérance sur 20 tours par proportionnalité	L'élève calcule l'espérance de la variable Y égale au nombre de cases avancées par un joueur sur 20 tours, à partir de l'espérance de la variable X définie dans la partie A, en multipliant l'espérance sur un tour (correctement ou non calculée) par 20.
calcul de l'espérance sur 20 tours : autre calcul	L'élève calcule l'espérance sur 1 tour, puis emploie divers artifices pour évaluer une espérance sur 20 tours, mais ne multiplie pas l'espérance de X par 20.
évaluation d'une « espérance » par le biais de représentativité	L'élève ne calcule pas l'espérance de la variable X ; il explique que sur les 20 tours, on doit retrouver le nombre de cases avancées (0, 1, 2 ou 3) en proportion correspondant aux probabilités calculées sur 1 tour, comme pour un très grand nombre d'épreuves. Nous avons trouvé cette solution chez des élèves qui avaient une loi de probabilité fausse, avec un dénominateur multiple de 10.

4.2. Caractères liés à la simulation

caractère	description de ce caractère
Lorsque l'élève s'appuie sur la simulation, combien de simulations propose-t-il de réaliser ? L'élève a pu indiquer une quantité, ou indiquer un ordre de grandeur par une locution :	
quantité	
une fois	
beaucoup, nombreuses	
plusieurs	
un grand nombre	
un très grand nombre de fois / de très nombreuses fois	
L'élève a pu proposer de faire une étude statistique des résultats obtenus grâce à la simulation. Le calcul de divers indicateurs statistiques a été proposé :	
la moyenne **ou la médiane (rare)**	En répétant un certain nombre de fois la simulation de 20 tours, on obtient une série statistique sur laquelle l'élève propose de calculer divers indicateurs : le plus souvent la moyenne
le nombre de fréquence max	l'élève propose de retenir le nombre de cases avancées qui a la plus grande fréquence
nombre de fréquence min	l'élève propose de retenir le nombre de cases avancées qui a la plus petite fréquence

4.3. Autres caractères

Certains élèves se sont appuyés à la fois sur la loi de probabilité déterminée dans la partie A. du problème, et sur la simulation d'un tour pour un joueur ; ils ont pu être amené à tenter de faire un lien entre les deux aspects : simulation et modélisation probabiliste.

Certains élèves ont tenté de prendre en compte le nombre de joueurs.

Enfin, nous avons essayé de déterminer dans les réponses quelques indices de la prise en compte (ou non) d'une incertitude quant à la possibilité de la partie de s'achever dans les temps impartis.

caractère	description de ce caractère
Les termes choisis par l'élève pour répondre à la question peuvent indiquer s'il a plus ou moins pris en compte le fait que la décision prise comporte une incertitude.	
certitude	Dans la formulation de sa conclusion, l'élève semble exprimer une certitude quand à ce que produira le choix qu'il propose : « on est sûr que », « devra », etc.
incertitude	D'une manière plus ou moins explicite, l'élève rend compte d'une incertitude quant à ce que produira le choix (« devrait », « espérer », etc.), ou présente son choix comme une valeur possible parmi d'autres (« environ », etc.).
Certains élèves essayent de tenir compte des fluctuations de la valeur d'une variable aléatoire autour de la valeur de son espérance ou de la moyenne observée, et proposent de prendre en compte ces fluctuations en donnant une marge d'erreur :	
marge d'erreur	L'élève propose de choisir de mettre sur le jeu un nombre de cases égal à l'espérance mathématique ou à la valeur moyenne observée, en ajoutant ou retranchant ce qu'ils considèrent comme une marge d'erreur acceptable. Un élève a tenté de calculer l'écart-type de la variable X de la partie A, sans savoir qu'en faire. Le choix de la valeur de la marge semble arbitraire.
cases en plus	l'élève ajoute des cases
cases en moins	l'élève enlève des cases
combien	en valeur algébrique.

La loi des grands nombres	
évocation de la loi des grands nombres	Certains élèves évoquent la loi des grands nombres.Cela peut être pour justifier qu'il faut répéter la simulation un très grand nombre de fois ; cela peut être pour indiquer le lien entre l'espérance mathématique et la moyenne observée pour des élèves qui se sont appuyés à la fois sur la modélisation probabiliste et sur la simulation.
lien moyenne / espérance	Certains élèves s'appuient en première intention sur la simulation, puis, ne pouvant donner de réponse numérique faute d'un passage sur machine pour mettre en œuvre la simulation, estiment la valeur moyenne du nombre S de cases avancées qu'ils auraient observé à l'aide de l'espérance de S, en indiquant le lien entre les deux nombres. Ce lien peut être correctement énoncé ou énoncé en révélant une conception erronée (par exemple l'égalité entre la valeur moyenne observée et l'espérance).
Le nombre de cases à mettre sur le jeu dépend du nombre de joueurs (2 ou 3 disaient les énoncés). L'importance du nombre de joueurs ne va pas de soi. Elle est d'autant plus difficile à prendre en compte que le jeu se joue à deux ou à trois joueurs. Un élève a buté sur le choix entre deux et trois joueurs, en concluant qu'on ne peut pas conseiller Monsieur Dupond. Tous les autres élèves qui ont essayé de prendre en compte le nombre de joueurs, ont raisonné avec trois joueurs.	
prise en compte de 3 joueurs	L'élève prend plus ou moins en compte la présence de trois joueurs. Le plus souvent cette prise en compte consiste à considérer qu'il y a proportionnalité entre le nombre de cases à mettre sur le jeu et le nombre de joueurs : on divise alors l'espérance du nombre Y de cases avancées sur 20 tours par un joueur (ou sa valeur moyenne) par 3.

4.4. Quelques précisions sur les caractères retenus

4.4.1. La détermination de l'espérance

Le plus souvent l'élève calcule l'espérance de la variable égale à l'écart absolu des points amenés lorsqu'on jette les deux dés : c'est l'espérance d'avancée pour un tour pour un joueur, puis il multiplie cette valeur par 20 (pour 20 tours).

Rappelons que

• L'espérance (théorique) pour un joueur qui joue 20 tours est de 25

• Pour une loi de probabilité erronée correspondant à des paires, l'espérance mathématique est de 20

Ce calcul utilise la linéarité de l'espérance : si X_i est le nombre de cases avancées au $i^{ème}$ tour par un joueur donné, le nombre de cases avancées sur 20 tours est $S = \sum X_i$, et $E(S) = \sum E(X_i)$; comme ici les X_i ont tous la même espérance : 1, 25, $E(S) = 20 \times 1{,}25 = 25$, mais les variables X_i sont définis sur l'univers produit des univers associés au lancer des deux dés pour chacun des 20 tours (nous avons détaillé le calcul dans l'analyse a priori).

Ce résultat n'est pas au programme de première, ni à celui de terminale, et si la plupart des élèves ont considéré qu'il allait de soi, quelques uns cependant ont semblés gênés pour calculer l'espérance de Y.

Nous doutons que ce soit le raisonnement évoqué ci-dessus qui a mené au calcul :

$E(S) = 20 \times E(X)$ où X est la variable aléatoire égale au nombre de cases avancées par un joueur sur un tour. Nous avons dit dans l'analyse a priori que probablement il y avait un jeu sur les mots : on espère avancer de 1,25 cases pour un tour, donc on espère avancer de :

$20 \times 1{,}25 = 25$ cases pour 20 tours.

Les réponses (voir leur transcription dans le document annexé) sont souvent formulées dans ces termes là. Aucun élève n'a introduit la variable aléatoire 20 X (qui aurait d'ailleurs posé un problème dans la définition des issues à considérer, cf. l'espace probabilisé produit évoqué plus haut).

Remarquons que les énoncés ne demandaient pas de calculer l'espérance d'une variable aléatoire, et que les élèves qui ont calculé l'espérance de S (variable définie ci-dessus), ont pris l'initiative de ce calcul, et ont vu qu'il passait par le calcul de l'espérance de la variable X dont la loi de probabilité avait été déterminée au A. Cela dépasse les initiatives des questions ouvertes du baccalauréat vues à ce jour.

Des élèves n'ont pas essayé de calculer une espérance de variable aléatoire, mais ont observé, que l'écart de plus forte probabilité était de 1 et ont organisé leur conseil autour de ce constat, proposant alors de prendre 20 cases.
Ils ont ajouté parfois des remarques « astucieuses » telles que : la probabilité d'avoir un écart de 0 ou 2 est de 1/2, « par conséquent » sur 2 tours le joueur avance de deux cases, ce qui fait une moyenne de 1 case par tour avec une probabilité de 1/2, donc cela renforce l'idée que sur un tour, un joueur avance environ d'une case, et donc de 20 cases sur 20 tours.

Certains élèves ont déterminé une *espérance* du nombre de cases avancées par un joueur sur 20 tours en utilisant le « biais de représentativité »
Nous avons observé cette démarche particulièrement chez des élèves qui avaient une loi de probabilité fausse, pour laquelle les probabilités des issues avaient un dénominateur divisant 20.
En particulier, la loi fausse associée à des paires, donne des probabilités des issues, qui, mises au dénominateur 20 (car il y a 20 tours), sont : 8/20, 6/20, 4/20 et 2/20.
Ces élèves ont donc d'abord ramené les probabilités calculées au dénominateur 20 et se sont ensuite attachés aux numérateurs : « les cas favorables » : sur 20 cas, il y en a 8 favorables à « avancer de 0 case », 6 favorables à « avancer d'une case », etc.
« Donc » sur 20 tours, le joueur avance 8 fois de 0 cases, 6 fois d'une case, 4 fois de 2 cases et 2 fois de 3 cases, ce qui fait au total : $0 + 6 + 4\times2 + 2\times3 = 20$ cases.
Nous citerons au chapitre 18 des réponses illustrant ce biais de représentativité.

4.4.2. La prise en compte de l'incertitude

Cette prise en compte, de façon plus ou moins nette, de l'incertitude dans la réponse d'un élève, nous paraît un indice qu'il a peut-être été, dans sa compréhension du problème posé, au delà de ce qui est habituellement attendu (le calcul d'une espérance ou d'une valeur moyenne observée), qu'il a réalisé, peut-être, qu'il faudrait tenir compte de la fluctuation autour de ces deux valeurs, ou que le choix de l'une de ces valeurs ne permet pas d'assurer que le contrat de durée sera respecté.

Nous avons ainsi rencontré (le numéro entre parenthèses est la référence de l'élève)
- « *Il faut mettre **environ** … cases* », (plusieurs élèves)
- « *avec … cases, le jeu se termine(ra) en 20 tours **en moyenne*** » (15)
- «*… le plateau **pourrait** être constitué de 20 cases, **sachant qu'un joueur peut également***

avancer de plus d'une case par tour » (16),

• « *une* ***quinzaine*** *de cases* » (7) ,

• « *il faut que le jeu ne soit* **pas plus long que** *l'espérance* » (45),

• « *pour que le jeu* ***puisse être souvent gagné****, le plateau de jeu doit faire … cases* » (50),

• on calcule la moyenne sur un grand nombre de simulations, « ***pour avoir une idée du nombre de case****s séparant la case de départ, de la case d'arrivée* »,

• « *Monsieur Dupond devrait donc faire un plateau de 25 cases pour* ***espérer*** *que la partie finisse en 30 minutes.* »,

• (130) essaye de tenir compte de la fluctuation autour de l'espérance calculée (30), et pour cela tente un calcul faux de l'écart-type (qu'il évalue à environ 2,23) « *ce qui est un peu élevé, on peut donc conseiller à M. Dupond de* ***mettre un peu moins de 30 cases s'il veut que le jeu dure 30 minutes ou moins de manière catégoriqu****e* »,

• (140) « ***En théori****e,(…), il faudrait que le nombre de cases* ***idéal*** *(…) soit 25 (…)* ».

Certains élèves semblent, au contraire, occulter le caractère aléatoire du jeu, ou tout au moins avoir mal compris que l'espérance d'une variable aléatoire, comme sa valeur moyenne observée, ne sont que des indicateurs autour desquels va fluctuer la variable aléatoire « nombre de cases avancées » lorsqu'on jouera :

• « *Le nombre de cases séparant la case de départ de la case d'arrivée* ***doit*** *être de 25 cases.* » (88)

• L'élève a calculé l'espérance du nombre de cases avancé par joueur sur 20 tours, et a trouvé 35. « *Pour* ***être certain*** *qu'il y aura un gagnant au bout de 20 tours (ou 30 min), il faudra donc que Monsieur Dupond choisisse un nombre de cases inférieur à 35* » (109)

• L'élève a calculé une espérance d'avancée sur 20 tours pour chaque joueur de 25 cases. « *Il doit donc y avoir au maximum 25 cases. Mais dans ce cas le jeu pourrait se terminer au 20^{e} tour. Avec 20 cases (donc 18 entre la case de départ et celle d'arrivée)* ***le jeu sera toujours terminé avant le dernier tour*** » (123).

Remarquons que certains élèves expriment à la fois de la certitude et de l'incertitude :

• « *Si le jeu est composé de 20 cases alors il y aura une* ***forte probabilité que tous les joueurs finissent la partie****. Il y aura alors certainement un joueur qui finira la partie avant la fin des 20 tours.* » (72)

4.4.3. Le lien entre les deux approches

Certains élèves se sont appuyés dans leur réponse à la question B. à la fois sur la loi de probabilité et sur la simulation demandées dans la partie A. Certains ont fait un lien entre la moyenne observée sur une répétition de la simulation, et l'espérance sur 20 tours ; ce lien a pu être correctement exprimé, mais également mal compris (ou mal exprimé) :

• l'espérance calculée est 25 ; l'élève ajoute : « *Sur un grand nombre de parties, les résultats des joueurs tendent vers 25 cases* » (50)

Il s'agit là de la confusion que nous avons évoquée au chapitre 2, paragraphe 5 : lorsqu'on répète un grand nombre de fois l'expérience, le hasard tendrait à s'estomper ; ce n'est plus la valeur moyenne qui tend à se stabiliser vers 25, mais le nombre de cases avancées par chacun des joueurs.

• L'élève propose que Monsieur Dupond fasse 10 simulations. « *Ensuite il calcule la moyenne de ces nombres et l'arrondit par défaut à l'unité près. Si on appelle S la somme en C21, il a calculé l'espérance de cette variable.* » (104)

L'élève a compris que dans la cellule C21 du tableur, il a une variable aléatoire S, mais n'a pas réalisé que d'une part la valeur moyenne des valeurs observées pour S sur 10 simulations est sujette à fluctuations, et que d'autre part sur 10 épreuves, cette moyenne n'a pas de raison d'être voisine de E(S), a fortiori d'être égale à E(S).

CHAPITRE 17

Analyse des réponses des élèves (partie A)

Note : nous avons fait dans ce chapitre plusieurs tests du χ^2 ; nous avons regroupé dans un tableau annexé à ce chapitre les valeurs des χ^2 observés, leurs probabilités et l'interprétation du test, nous contentant dans ce chapitre d'indiquer si les écarts observés sont ou non significatifs. <u>*Le seuil choisi a toujours été 5 %.*</u>

1. Les réponses regroupées par catégories

Nous avons croisé les réponses aux deux questions de la partie A classées selon les groupes définis précédemment. Les résultats sont regroupés dans le tableau suivant :

modélisation / simulation	exacte	paires	eq[03]	8 doubles	Cas favorables	néant	Autre	total
exacte	70	9	0	2	2	0	0	83
incomplet	14	1	0	0	0	0	0	15
eq[03]	7	1	4	0	0	0	0	12
lancer	3	1	0	0	0	1	0	5
néant	10	2	0	0	1	0	1	14
Autres	4	0	0	0	0	0	1	5
total	108	14	4	2	3	1	2	134

<u>tableau 1</u>

Réponses croisées, classées par groupes, aux deux questions « simulation » et «loi de probabilité » <u>pour toute la population expérimentale</u>.

Pour la simulation, nous avons regroupé les groupes : « **loi de probabilité »**, « **imitation** », et « **Dé5** » sous la rubrique « **exacte »** ; les groupes **« imitation incomplète** » , « **imitation des lancers** » et « **loi équiréparties sur {0 ; 1 ; 2 ; 3}** » sont ici codés « **incomplet** », « **lancer** » et « **eq[03]** ».

Pour la loi de probabilité, nous avons codé les réponses correctes par « **exacte** », les groupes sont dans le tableau dans la colonne étiquetée par leurs noms, à l'exception des groupes « **nombre de cas favorables** » et **« loi équirépartie sur [0 : 3]** » codés ici « **Cas**

favorables » et **«eq[03]** ».

Nous avons également croisé les groupes de réponses aux deux questions « simulation » et « loi de probabilité » selon :

• les modalités (énoncé 1 ou énoncé 2) dans les tableaux 2 et 3,

• les niveaux (terminale ou première) dans les tableaux 3 et 4.

modélisation / simulation	exacte	paires	eq[03]	8 doubles	Cas favorables	néant	Autre	total
exacte	39	6	0	0	0	0	0	45
incomplet	5	0	0	0	0	0	0	5
eq[03]	2	0	3	0	0	0	0	5
lancer	1	1	0	0	0	0	0	2
néant	6	1	0	0	0	0	0	7
Autres	1	0	0	0	0	0	1	2
total	54	8	3	0	0	0	1	66

tableau 2

Réponses croisées, classées par groupes, aux deux questions « simulation » et «loi de probabilité » données par les élèves ayant reçu l'énoncé 1.

modélisation / simulation	exacte	paires	eq[03]	8 doubles	Cas favorables	néant	Autre	total
exacte	31	3	0	2	2	0	0	38
incomplet	9	1	0	0	0	0	0	10
eq[03]	5	1	1	0	0	0	0	7
lancer	2	0	0	0	0	1	0	3
néant	4	1	0	0	1	0	1	7
Autres	3	0	0	0	0	0	0	3
total	54	6	1	2	3	1	1	68

tableau 3

Réponses croisées, classées par groupes, aux deux questions « simulation » et «loi de probabilité » données par les élèves ayant reçu l'énoncé 2.

modélisation / simulation	exacte	paires	eq[03]	8 doubles	Cas favorables	néant	Autre	total
exacte	40	6		1				47
incomplet	6	1						7
eq[03]	2		2					4
lancer						1		1
néant	5	1						6
Autres	1						1	2
total	54	8	2	1	0	1	1	67

tableau 4

Réponses croisées, classées par groupes, aux deux questions « simulation » et « loi de probabilité » données par les élèves de terminale.

modélisation / simulation	exacte	paires	eq[03]	8 doubles	Cas favorables	néant	Autre	total
exacte	30	3		1	2			36
incomplet	8							8
eq[03]	5	1	2					8
lancer	3	1						4
néant	5	1			1		1	8
Autres	3							3
total	54	6	2	1	3	0	1	67

tableau 5

Réponses croisées, classées par groupes, aux deux questions « simulation » et « loi de probabilité » données par les élèves de première.

1.1. Les réponses pour la simulation

Le choix de la technique

Rappelons que deux techniques étaient attendues : *imitation au plus près* ou *simulation de la loi de probabilité*.

Parmi les élèves qui ont choisi la technique d'*imitation au plus près*, certains n'ont pas complètement menée à bien la simulation, oubliant de faire calculer le nombre total de cases avancées à l'issue des 20 tours

technique	*imitation au plus près*		*simulation de la loi de probabilité*	autres réponses : eq[03]+lancer + néant+autres	total
	menée à bien	incomplète			
effectif	82	15	1	36	134
pourcentage	61 %	11 %	< 1 %	27 %	100

tableau 6

La technique choisie par les élèves pour simuler 20 tours pour un joueur.

72 % des élèves ont donc choisi de simuler 20 tours pour un joueur par la technique d'*imitation au plus près*.

Un seul élève a choisi de simuler la loi de probabilité de l'expérience aléatoire. Sa simulation est exacte, et correspond à la simulation d'un tirage d'une urne contenant 16 éléments.

Cet élève a composé avec l'énoncé 1 : il venait donc de déterminer la loi de probabilité de la variable X égale au nombre de cases avancées sur un tour par un joueur, avant d'aborder la deuxième question de son énoncé, qui lui demandait d'établir une feuille de calcul pour simuler 20 tours.

La qualité des réponses

83 élèves ont produit une simulation exacte et 15 une simulation « presque exacte » (élèves ayant oublié de cumuler le nombre de cases avancées).

Nous avons établi une comparaison de la qualité des réponses pour la simulation

• selon l'énoncé (tableau 7)

• et selon le niveau (tableau 8)

simulation proposée / modalité	*imitation au plus près* ou *simulation de la loi de probabilité*		eq[03]+lancer + autres	abstention	total
	exacte	presque exacte			
Enoncé 1	45	5	9	7	66
Enoncé 2	38	10	13	7	68
total	83	15	22	14	134

tableau 7

Réponses proposées pour la simulation, selon les énoncées

simulation proposée / niveau	*imitation au plus près* ou *simulation de la loi de probabilité*		eq[03]+lancer + autres	abstention	total
	exacte	presque exacte			
Terminale	47	7	7	6	67
Première	36	8	15	8	67
total	83	15	22	14	134

tableau 8

Réponses proposées pour la simulation, selon les niveaux

Le test du chi2 n'indique pas, au risque 5%, de différence significative pour les écarts observés dans la répartition entre les quatre groupes de réponses données pour la simulation, que ce soit les écarts selon les énoncés, que ce soit les écarts selon les niveaux.

1.2. Les réponses pour la loi de probabilité

108 élèves ont déterminé correctement la loi de probabilité de la variable aléatoire X égale au nombre de cases avancées par un joueur sur un tour.

Remarquons que le taux de réussite à la détermination de la loi de probabilité (80,6%) est supérieur au taux de réussite pour la simulation (61,9%, et 73,1% si l'on prend en compte les réponses incomplètes, c'est à dire les réponses pour lesquelles le nombre total de cases avancées n'a pas été calculé).

La détermination de la loi de probabilité pour la somme des points amenés par le lancer de deux dés cubiques équilibrés est un exercice de référence, en principe travaillé par tous les élèves. Nous avons indiqué au chapitre 15 que cet exercice a été fait en classe dans la progression du cours de probabilité mené par le professeur pour tous les élèves de première concernés par notre expérimentation ; cet exercice a fait partie des révisions du programme de première S pour les élèves de terminale.

Nous avons également fait, pour la détermination de la loi de probabilité, une comparaison selon les énoncés et selon les niveaux (tableaux 9 et 10).

réponse : / modalité	loi de probabilité correcte	autres réponses	total
énoncé 1	54	12	66
énoncé 2	54	14	68
total	108	26	134

tableau 9

Réponses proposées pour la loi de probabilité, selon les énoncées.

Il n'y a pas de différence significative dans la qualité des réponses selon les énoncés.

Un écart significatif en faveur de la deuxième modalité (la détermination de la loi de probabilité après la simulation de l'expérience aléatoire) aurait pu être un argument pour justifier l'ordre des questions que nous avons pu observer dans les sujets des épreuves pratiques expérimentales (chapitre 3, paragraphe 1, et annexes du chapitre 3) : on aurait pu avancer l'hypothèse qu'une *imitation au plus près* mise en œuvre en simulant le lancer des

deux dés dans deux cellules distinctes prépare la détermination de l'univers des *issues physiques*, ensemble des couples (x ; y) d'éléments de l'ensemble {1 ; 2 ; 3 ; 4}.
Dans les résultats observés ici, l'ordre des tâches (simulation et modélisation probabiliste) semble neutre par rapport à la qualité de la modélisation probabiliste.

Et enfin le tableau 10 des réponses selon les niveaux :

réponse : / niveau	loi de probabilité correcte	autres réponses	total
Terminale	54	13	67
Première	54	13	67
total	108	26	134

tableau 10

Réponses proposées pour la loi de probabilité, selon les niveaux.

1.3. Les réponses croisées aux deux questions (simulation et loi de probabilité)

108 élèves ont correctement déterminé la loi de probabilité, et seulement 83 ont une simulation correcte.
Si pour la simulation on réunit les réponses exactes et « presque exactes » (élèves n'ayant pas cumulé les cases avancées), ce sont 95 élèves qui ont correctement calculé l'écart des points amenés.

Notons que 10 élèves ont correctement déterminé la loi de probabilité, mais n'ont pas abordé la simulation.

Précisons les réponses selon les énoncés (tableau 11) et selon les niveaux (tableau 12) :

réponses / modalité	exacte pour la loi de probabilité		exacte pour la loi de probabilité <u>et</u> abstention ou réponse inexacte pour la simulation	abstention ou réponse inexacte pour la loi de probabilité <u>et</u> exacte ou presque exacte pour la simulation	abstention ou réponse inexacte pour la loi de probabilité <u>et</u> abstention ou réponse inexacte pour la simulation	total
	<u>et</u> exacte pour la simulation	**<u>e</u>**t presque exacte pour la simulation				
énoncé 1	39	5	10	6	6	66
énoncé 2	31	9	14	8	6	68
total	70	14	24	14	12	134

<u>tableau 11</u>

Réponses croisées aux deux questions (simulation et loi de probabilité) selon les énoncés

Les écarts observés entre les deux modalités ne sont pas significatifs.

réponses / niveau	exacte pour la loi de probabilité		exacte pour la loi de probabilité **et** abstention ou réponse inexacte pour la simulation	abstention ou réponse inexacte pour la loi de probabilité **et** exacte ou presque exacte pour la simulation	abstention ou réponse inexacte pour la loi de probabilité **et** abstention ou réponse inexacte pour la simulation	total
	et exacte pour la simulation	et presque exacte pour la simulation				
terminale	40	6	8	8	5	67
première	30	8	16	6	7	67
total	70	14	24	14	12	134

<u>tableau 12</u>

Réponses croisées aux deux questions (simulation et loi de probabilité) selon les niveaux

Les écarts observés selon les niveaux ne sont pas significatifs.

1.4. Les hypothèses de modélisation

La traduction des hypothèses de modélisation dans chacune des deux tâches

Nous allons reprendre les données du **tableau 1** pour essayer d'analyser les réponses du point de vue des hypothèses de modélisation. Pour cela, nous avons regroupé certaines classes, pour lesquelles nous pouvons dire qu'il y a un même choix d'hypothèses de modélisation.

Pour la simulation, nous pouvons rassembler les trois groupes : « exacte », « incomplet », « lancer », pour lesquels il apparaît que les hypothèses de modélisation, données dans l'énoncé par la phrase « *Dans ce jeu, le joueur dont c'est le tour lance deux dés tétraédriques équilibrés* », sont traduites par deux appels de « Alea.entre.bornes », que ce soit Alea.entre.bornes(1 ; 4) ou Alea.entre.bornes(0 ; 3) [67], ou Alea.entre.bornes(1 ; 6) pour les élèves qui ont travaillé avec des dés cubiques.
Nous regroupons ces réponses sous la rubrique « correcte ou presque correcte » pour les hypothèses de modélisation.

Pour la détermination de la loi de probabilité, nous rassemblons, conformément à l'analyse que nous avons faite au chapitre 14, les groupes « correcte », « paire », « 8 doubles » et « cas favorables » qui correspondent à une prise en compte correcte des hypothèses de modélisation sur les deux dés, même si le dénombrement des issues possibles peut être erroné.
Nous regroupons ces réponses sous la rubrique « correcte ou presque correcte » pour les hypothèses de modélisation.

Pour les deux tâches (loi de probabilité et simulation), nous avons un groupe codé dans les tableaux 1 à 5 du début de ce chapitre par « eq[03] » : les élèves de ces groupes ont choisi la loi équirépartie sur les 4 *issues intéressantes*, c'est à dire les 4 valeurs possibles pour la variable aléatoire X ; ces groupes correspondent, comme nous l'avons analysé au chapitre 14 paragraphe 1, au biais d'équiprobabilité.

Enfin, nous avons rassemblé les groupes « néant » et « autre » des tableaux 1 à 5.

Ces regroupement étant faits, nous obtenons les groupes « correct ou presque correct », « écarts équiprobables », « autre réponse » pour les deux questions (simulation et loi de

[67] Quand il y a deux appels, et que la différence des nombres sortis apparaît dans le tableur.

probabilité) :

modélisation / simulation	correct ou presque correct	écarts équiprobables	autre réponse	total
correct ou presque correct	102	0	1	103
écarts équiprobables	8	4	0	12
autre réponse	17	0	2	19
total	127	4	3	134

tableau 13

Traduction des hypothèses de modélisation : réponses croisées pour les deux tâches

Il ressort de ce tableau que 103 élèves ont correctement traduit, dans leur simulation, les hypothèses de modélisation sur chacun des deux dés, et 127 élèves ont correctement traduit ces hypothèses dans la détermination de la loi de probabilité.

Nous voyons qu'il y a une moindre réussite, de ce point de vue, pour la simulation que pour la loi de probabilité.

Le biais de la loi équirépartie

Nous avons présenté dans le chapitre 9 le modèle que nous proposons pour les liens entre l'expérience aléatoire, son modèle probabiliste et la simulation que l'on propose de cette expérience.

Nous avons montré que la simulation peut être une *imitation au plus près*, ce qui, à une exception près, a été le cas pour ceux des élèves qui ont produit une simulation correcte ou incomplète.

Nous avons indiqué qu'il y a une étape commune aux deux tâches de modélisation probabiliste et simulation par *imitation au plus près*, étape que nous avons appelée « étape de pré-modélisation », où sont choisies les *hypothèses de modélisation*.

Dans notre expérimentation, ces hypothèses de modélisation (les dés sont équilibrées, donc pour chaque dé, les quatre issues sont équiprobables) étaient données dans l'énoncé, donc n'étaient pas à la charge de l'élève, à cela près qu'il lui incombait de traduire ces hypothèses de modélisation « en actes », dans chacune des deux tâches.

12 élèves ont choisi de simuler les *événements intéressants* (les valeurs des écarts : 0, 1, 2 et 3) en les considérant comme équiprobables. Cette erreur est analysée comme une manifestation du « biais d'équiprobabilité » décrit au chapitre 2, paragraphe 4 et au chapitre 14 paragraphe 1.
Ces 12 élèves ont donc une étape de pré-modélisation incorrecte.

Ces 12 élèves ont commis l'erreur au moins pour la simulation, mais parmi ces 12 élèves, seulement 4 ont choisi la loi équirépartie sur l'ensemble {0 ; 1 ; 2 ; 3} et pour la simulation, et pour la loi de probabilité de la variable aléatoire X, faisant ainsi preuve d'une certaine cohérence.
Rappelons que avons souligné dans le chapitre 16, paragraphe 1 que le classement de deux de ces élèves dans les deux groupes « loi équirépartie » était discutable.

Des choix discordants
En revanche, parmi ces 12 élèves, 7 élèves ont une loi de probabilité de la variable X correcte, et un élève est classé dans le groupe « paires ». Ce sont donc 8 élèves qui n'ont pas fait le même choix d'hypothèses de modélisation pour les deux tâches.
Il semble donc que pour ces 8 élèves, les analyses pour chacune des deux tâches de simulation et de détermination de la loi de probabilité ont été faites de manières disjointes.
Les effectifs sont faibles, donc la comparaison selon les énoncés ou selon les classes est délicate ; remarquons que l'on trouve ces discordances pour les deux modalités, et pour les deux niveaux.

CHAPITRE 18

Analyse des réponses (partie B)

1. Modélisation ou simulation ?

Concernant les éléments sur lesquels les élèves se sont appuyés pour répondre à la question B. (« comment conseiller M. Dupond »), nous avons défini, au paragraphe 1.4. du chapitre 16, différents groupes de réponses : « **modélisation exclusivement** », « **simulation exclusivement** », « **modélisation et simulation** », « **néant** » et « **autre** ». Ce dernier groupe ne comporte aucune réponse comportant un argument pertinent qui aurait pu justifier la création d'un groupe particulier.

Nous avons comparé la distribution des réponses selon les énoncés (tableau 1) :

la réponse s'appuie sur / modalité	la modélisation seule	la simulation seule	la simulation et la modélisation	autre réponse	abstention	total
énoncé 1	14	24	4	12	12	66
énoncé 2	36	5	1	14	12	68
total	50	29	5	26	24	134

tableau 1

Les réponses à la question B. (« comment conseiller ») selon les énoncés.

La valeur du χ^2 est de 24, 058 avec une probabilité inférieur à 0,0001

Plus de 80 % des effectifs théoriques dépassent 5, et l'effectif total dépasse 40. Nous avons cependant effectué le test avec la correction de Yates : le χ^2 corrigé vaut 20,817 ; sa probabilité est également inférieure à 10^{-4}.

Selon le test du χ^2, les écarts observés dans la distributions des réponses selon les énoncés sont très significatifs.

Pour préciser le résultat, nous avons d'abord opposé les réponses du groupe « modélisation seule » à tous les autres groupes (tableau 2), puis les réponses du groupe « simulation seule » à tous les autres groupes (tableau 3) :

la réponse s'appuie sur / niveau	la modélisation seule	toutes les autres réponses	total
énoncé 1	14	52	66
énoncé 2	36	32	68
total	50	84	134

<u>tableau 2</u>

Influence de l'énoncé dans le choix d'une argumentation à la question B. s'appuyant sur la seule modélisation probabiliste.

Le test du χ^2 indique que les écarts observés dans la distribution des réponses selon les énoncés sont très significatifs (χ^2 = 14,415, probabilité inférieur à 0,0002).

Le fait d'avoir déterminé la loi de probabilité de la variable aléatoire X dans la question qui précédait immédiatement la question B (énoncé 2) semble avoir eu une influence sur le choix de l'argument, en faveur de la modélisation probabiliste.

la réponse s'appuie sur / modalité	la simulation seule	toutes les autres réponses	total
énoncé 1	24	42	66
énoncé 2	5	63	68
total	29	105	134

<u>tableau 3</u>

Influence de l'énoncé dans le choix d'une argumentation à la question B. s'appuyant sur la seule simulation.

Le test du χ^2 indique encore des écarts très significatifs dans la distribution des réponses selon les énoncés (χ^2 = 16,622, probabilité inférieure à 0,0001).

Le fait d'avoir déterminé la simulation dans la question qui précédait immédiatement la question B (énoncé 1) semble avoir eu une influence sur le choix de l'argument, en faveur de la simulation.

Il semble donc que la question qui précédait immédiatement la question B. du problème a eu une certaine influence sur le raisonnement tenu par les élèves pour répondre à cette question B. Ce constat va dans le sens de Maury (1986) sur l'influence possible de la formulation de la question dans les réponses données par les élèves (cf. chapitre 13, paragraphe 2.3).

Nous avons également comparé la distribution des réponses selon les niveaux (tableau 4) :

la réponse s'appuie sur / niveau	la modélisation seule	la simulation seule	la simulation et la modélisation	autre réponse	abstention	total
terminale	23	20	3	10	11	67
première	27	9	2	16	13	67
total	50	29	5	26	24	134

tableau 4

Les réponses à la question B. (« comment conseiller ») selon les niveaux.

Le test du χ^2 n'indique pas d'écart significatif dans la distribution des réponses selon les niveaux (χ^2 = 6,244, probabilité de 0,182).

Au delà delà de la comparaison des distributions des réponses selon les énoncés ou selon les niveaux, il faut remarquer que sur l'ensemble de l'échantillon, les élèves se sont plus appuyés sur la modélisation que sur la simulation : 50 élèves sur les 84 élèves (soit 59,5 %) qui ont explicitement utilisé la partie A. du problème (détermination de la loi de probabilité de X et simulation de 20 tours pour un joueur), ont utilisé exclusivement la détermination de la loi de probabilité X, contre 29 élèves sur les 84 (34,5 %) qui ont utilisé uniquement la simulation [68] . Il est vrai qu'en l'absence d'un passage sur ordinateur, cette simulation est restée théorique puisqu'elle n'a pas produit de résultat. Cependant la question telle qu'elle était posée

[68] 6 % des élèves se sont appuyés à la fois sur la modélisation probabiliste et la simulation.

(« comment conseiller ») appelait dans la réponse une méthode plus qu'un résultat déterminé après un passage sur ordinateur.

Il semble utile de re-situer les tâches de *simulation* et *modélisation probabiliste* dans ce qui paraît être encore le contrat didactique (implicite, autant pour les élèves que pour les enseignants) : la norme pourrions-nous dire, est le modèle probabiliste, c'est à son acquisition que visent les programmes.
La simulation est présentée comme une approche expérimentale, pour aller à cette norme.
Nous allons nous permettre un rapprochement avec une autre notion du programme des classes de première et terminale S : *les lieux géométriques*.
La démarche préconisée (autant que possible) est de conjecturer un résultat à partir de la construction du lieu à l'aide d'un logiciel dynamique de géométrie (géogébra, géoplanW, cabri-géomètre, etc.) puis de démontrer que le lieu est bien ce que l'on a conjecturé. Le contrat, là, est explicite : ce qui est attendu, c'est la « preuve mathématique ».
De manière générale, l'utilisation de l'outil informatique en classe de mathématiques permet de conjecturer un résultat qu'il faut (toujours selon le contrat) ensuite démontrer.
L'apport de l'outil informatique est donc peut-être compris alors comme une étape qu'on pourrait ensuite effacer pour passer aux choses sérieuses : faire des mathématiques « en démontrant ».

Sous cette hypothèse, il n'est peut-être pas surprenant que les élèves, dans le cadre d'un travail scolaire, aient essayé de répondre à la question B. (« comment aider M. Dupond »), pour certains d'entre eux, en utilisant plutôt une démarche qui paraît satisfaire le contrat, ou dont ils ont peut-être plus l'habitude.

2. Analyse des procédures proposées par les élèves pour aider M. Dupond

Il s'agit des procédures mises en œuvre par les 85 élèves classés dans au moins l'un des groupes « **modélisation** » ou « **simulation** », procédures présentées dans le chapitre 16, paragraphe 2.
Nous avons vu que ces procédures s'appuient majoritairement sur la modélisation probabiliste, quelques élèves, 5, s'appuyant sur les simulations et la modélisation probabiliste.

Nous présentons ci-dessous un tableau récapitulant les effectifs de chacun des caractères retenus pour décrire les réponses à la partie B.

caractère	**effectif**
prise en compte de 3 joueurs	5
le conseil s'appuie sur la modélisation	55
calcul de l'espérance sur 20 tours par proportionnalité	32
élèves ayant eu un problème avec le calcul de l' espérance	5
« biais de représentativité »	5
conseils s'appuyant sur la simulation	34
la simulation est répétée : une fois	0
beaucoup, de nombreuses fois	3
plusieurs fois	9
un grand nombre de fois	8
un très grand nombre / de très nombreuses fois	3
L'élève propose de calculer un indicateur statistique tel que : la moyenne ou la médiane (rare)	26
le nombre de cases de fréquence maximale	3
le nombre de cases de fréquence minimale	1
l'élève évoque un lien entre la moyenne observée et l' espérance mathématique	6
l'élève évoque la loi des grands nombres	4
la réponse de l'élève évoque une certitude quant aux conséquences de la décision	5
la réponse évoque une incertitude	35
une marge d'erreur est proposée	7
avec des cases en plus	2
avec des cases en moins	5

tableau 6

Effectifs des caractères étudiés dans les réponses à la question B. (« comment aider M. Dupond »).

3. Remarques complémentaires pour les réponses s'appuyant sur la modélisation

<u>3.1. Le calcul de l'espérance a posé un problème à 5 élèves</u>

Ces élèves ont annoncé un calcul de l'espérance de la variable X, nombre de cases avancées par un joueur à l'issue de un tour, mais n'ont pas conclu que l'espérance de S, nombre de cases avancées sur les 20 tours est égale à 20 × E(X). Citons l'élève (83) :

> *« Il faut calculer l'espérance de X*
>
> *E(X) = 0×1/4+1×3/8+2×1/4+3×1/8*
>
> *= 3/8+4/8+3/8*
>
> *= 10/8*
>
> *= 1,25*
>
> *On peut espérer avancer d'une case par tour, donc M. Dupond devrait faire un plateau composé de 19 cases séparant le départ de l'arrivée pour qu'au 20ème tour le pion arrive sur la case arrivé. »*

Et l'élève (86) :

> *« E(X) = 6/16 + 8/16 + 6/16*
>
> *= 20/16*
>
> *Au bout de 16 tours, le joueur avancera de 20 cases environ puis pour les quatre derniers tours (le joueur avancera d'environ 4 tours.)*
>
> *Le jeu devrait être constitué de 20 (à 25) cases. »*

Cet élève a donné les fractions de la loi de probabilité de la partie **A.** aux dénominateurs 16. Il semble qu'il applique la proportionnalité aux 16 premiers tours, puis considère que pour les quatre derniers tours, il convient d'espérer une avancée de 1 case par tour.

Pour l'élève (118) :

> « $E = 0 \times \frac{1}{4} + \frac{3}{8} \times 1 + 2 \times \frac{1}{4} + \frac{1}{8} \times 3 = \frac{3+3+4}{8} = \frac{10}{8} = \frac{5}{4} = 1{,}25$
>
> *Comme l'espérance est de 1,25 le joueur a plus de chance de faire 1 au dé.*
>
> *La case de départ ne compte pas comme une case : c'est le point 0*

Pour atteindre la case d'arrivée, qui compte comme une case, en 20 tours le jeu doit comporter 21 cases où la case 21 est l'arrivée et la 1 le départ. »
Car le joueur risque d'avancer d'1 case par tour donc de 20 cases en 20 tours.

L'élève (141) commence par s'appuyer sur les simulations ; on peut penser que pour avoir une réponse numérique à proposer, il a fait le lien entre la valeur moyenne et l'espérance, dont le calcul ne nécessite pas mise en œuvre effective de la simulation :

« Monsieur Dupond répète de nombreuses fois l'opération (1000 fois) il fait la moyenne des sommes obtenues. Cette moyenne correspondra au nombre de cases possibles pour espérer qu'il y ai un gagnant.
Sur les trois joueurs, un atteindra avant les autres cette moyenne au bout de 20 tours ou légèrement plus.
Cette moyenne correspondra à l'espérance mathématique de X.
$E(X) = 0 \times 4/16 + 6/16 + 2\times4/16 + 3\times2/16 = 20/16 = 5/4 \sim 1.$
On peut donc dire que Mr Dupond peut donner à son jeu 20 cases de la case de départ à celle d'arrivée. »

Et enfin l'élève (148) :

« On calcul l'espérance de X, tel que X représente le nombre moyen de cases parcourus pour chacun des joueurs en l'espace de 10 lancés donc de 10 tours :
$E(X) = \sum x_i\, p_i$
$E(X) = 0 \times 4/10 + 1 \times 3/10 + 2 \times 2/10 + 3/10$
$E(X) = 10/10 = 1$
Soit en 10 tours en moyenne chaque joueur avancerais de 1 case, je conseillerai donc a M. Dupond de limiter le nombre de cases à 2 si il ne veut pas dépasser la limite de 30 minutes. Cependant je dirai aussi a M.Dupond que plus de tours est important, plus il y a de chance que ces calculs soit proche de la réalité. Etant donné que le nombre de 20 tours au plus est imposé je lui dirai que il est probable d'avoir dans la réalité des résultats différents des miens ».

Dans cette réponse, l'interprétation même du calcul faux de E(X) = 10/10 est défaillante (encore peut-il s'agir au départ d'une grossière étourderie) : l'espérance calculée dit en réalité que l'espérance sur 20 tours est de 20 cases. Il semble que l'élève ait éprouvé quelques doutes quant à la pertinence de son résultat, ce qui l'amène à évoquer les écarts entre les calculs (sous-entendu théoriques) et la réalité, non pas en terme de fluctuations autour de la valeur théorique, mais en terme de résultats de calcul ayant plus de chances d'être proches de la réalité (observée ?) s'il y a plus de tours (référence probable à la loi des grands nombres).

Voici donc cinq façons plus ou moins astucieuses d'exploiter l'espérance mathématique calculée sur un tour, sans mettre en œuvre la propriété de proportionnalité, pour évaluer l'espérance sur 20 tours.
Quelques élèves ont choisi un tout autre angle d'attaque que le calcul d'une espérance.

3.2. Le biais de représentativité

Parmi ceux-ci, 5 élèves ont « succombé » au biais de représentativité évoqué dans notre première partie.

Illustrons cette procédure avec l'élève (80) :
Les probabilités des écarts 0, 1, 2 et 3 viennent d'être calculées : 4/10, 3/10, 2/10 et 1/10. Les valeurs sont fausses, mais le dénominateur 10 divise 20 : l'élève ramène tout au dénominateur 20, correspondant au nombre de tours, puis :

« 8/20 ; 6/20 ; 4/20 ; 2/20 → on sait qu'il y a 20 tours d'où le 20.

Pendant 8 tours, le joueur n'avance pas.	*0 cases*
Pendant 6 tours, le joueur avance d'une case	*6 cases*
Pendant 4 tours, le joueur avance de deux cases	*8 cases + 6 cases = 14 cases*
Pendant 2 tours, le joueur avance de trois cases	*6 cases + 14 cases = 20 cases*
Il faudrait donc 20 cases pour ce jeu, soit une case par tour. »	

L'élève (106) a trouvé pour loi de probabilité (1/4 ; 3/8 ; 1/4 ; 1/8) :

« Sur 20 tour un joueur obtient environ 8 fois le nombre 1 ($\frac{3}{8} \times 20 = 7{,}5$) ; 5 fois le

nombre 0 ; 5 fois le nombre 2 et 2 fois le nombre 3 (20 × $\frac{1}{8}$ = 2,5). Soit un joueur avance d'environ 8 + 5×2 + 3×2 = 22 cases. Le plateau doit contenir environs 22 cases. »

L'élève (133) a trouvé pour loi de probabilité (2/5 ; 3/10 ; 2/10 ; 1/10), et écrit :

> *« Selon la loi des grands nombres, sur 20 tours un joueur ne devrait pas avancer à 8 reprise, avancer de 1 case à 6 reprise, avancer de 2 cases à 4 reprises, avancer de 3 cases à 1 reprise. Donc en 30 minutes, un joueur devrait avancer de 20 cases. La partie ne devrait pas durer plus de 30 minutes, et le jeu se jouant à 2 ou 3 joueurs, Mr Dupond devrait mettre 7 cases à son jeu pour espérer avoir un gagnant dans les délais impartis. »*

L'élève (135) a pour loi de probabilité (4/8 ; 3/8 ; 2/8 ; 1/8). Il écrit :

> *« Etant donné que l'on connaît la probabilité pour chacune des issues on peut déterminer le nombre de case que devrait environ parcourir un joueur pour 20 tours. On va donc multiplier le numérateur de la probabilité de X auquel il correspond et faire la somme de ces résultat pour tous les X, Cet a dire :*
> *8×0 + 1×6 + 2×4 + 3×2 = 28.*
> *Pour une partie de 20 tours les joueurs devront avancer de plus ou moins 28 cases. C'est donc le nombre que l'on va conseiller à monsieur Dupond. Car de temps à autre des joueurs feront plus de 28 cases ou moins de 28 . »*

La loi de probabilité de cet élève faisait apparaître à l'origine des fractions de dénominateur égal à 20, ce qui explique peut-être ce raisonnement qui paraît, sans cette information, incompréhensible.

13 élèves sur les 55 qui se sont appuyés sur la modélisation pour conseiller Monsieur Dupond ont imaginé d'autres procédures.

3.3. Les autres procédures

Les autres propositions paraissent variées, mais reposent essentiellement sur l'observation des probabilités de chacune des issues (écart égal à 0, 1, 2 ou 3) pour un tour : c'est l'écart égal à 1 qui est le plus probable. Les élèves font alors des variations sur ce thème.

L'élève (7) :

> *« • p(X=1) est la plus importante. p(X=1) =* $\frac{6}{16}$
>
> *• p(X=0) = p(X=2)*
>
> *Le joueur a* $\frac{1}{4}$ *de chances de faire d'avancer de 2 cases d'un coup et* $\frac{1}{4}$ *de chance de ne pas avancer du tout, ce qui ferait une moyenne d'un avancement de 1 case sur deux tours.*
>
> *• p(X=3) =* $\frac{2}{16}$ *Il est donc peu probable d'avancer de 3 cases d'un coup.*
>
> *• Si jamais le joueur ne fait que des lancés l'amenant à avancer de 3 cases à la fois, il faudrait 20 ×3 = 60 cases.*
>
> *• Si jamais le joueur ne fait que des lancés l'amenant à avancer d'1 case à la fois, il faudrait 20 cases.*
>
> *Mais il y a aussi plus de probabilité d'arriver à avancer une fois de 0 cases et une autre fois de 2 cases p(0+2)=*$\frac{1}{2}$*. Donc d'avancer de 1 case sur 2 tours. Soit 10 cases au bout de 20 tours Il faudrait donc une quinzaine de cases. »*

En plus succinct, l'élève (10) :

> *« Si je pouvais aider M. Dupond à déterminer le nombre de cases nécessaires pour son jeu, je lui conseillerai que, puisqu'il veut qu'une partie ne dure surtout pas plus de 20 tours, il observe les probabilités ci-dessus et qu'il constate, que la probabilité la plus forte (mis à part le cas des doubles où le joueur n'avancerait pas) est pour le cas ou le joueur X n'avance que d'une seule case par tour. Soit 1×20 = 20 cases en tout, soit 18 cases entre la première et la dernière. »*

Ou l'élève (11), qui considère que les écarts 0 et 2 de même probabilité s'équilibrent pour une avancée de 1 :

« Nous pourrions conseiller à Mr Dupond de ne prendre qu'en considération le fait qu'un joueur tombe majoritairement sur des dés ayant un écart de 1. (puisque p(0) et p(2) sont égales mais lorsque le joueur obtient 0, n'avance pas du tout alors qu'avec 2, il avance 2 fois plus qu'en 1) »

L'élève (16) :

« Un joueur a une plus grande probabilité d'avancer d'1 case (selon A.2.) par tour
Une partie étant constituée de 20 tours, le joueur aura plus de chance, en tout, d'avancer de 20 cases.
Le plateau pourrait être constitué de 20 cases, sachant qu'un joueur peut également avancer de plus d'1 case par tour. »

Les 11 réponses reposent toutes sur ce constat que l'écart le plus probable est 1.

4. Remarques complémentaires pour les réponses s'appuyant sur la simulation

Les procédures s'appuyant sur la simulation concernent 35 élèves (dont cinq ont également évoqué la modélisation probabiliste).
La plupart de ces 35 élèves proposent de répéter la simulation de 20 tours pour un joueur, et de repérer un indicateur statistique sur la série statistique ainsi générée (ils ne formulent pas leur démarche dans ces termes).

<u>Le nombre de répétitions de la simulation</u>

Le nombre de simulations à effectuer n'est précisé que 10 fois :

nombre de simulations proposé dans la réponse	**> 1000**	**1000**	**100**	**10**
effectif de ces réponses	2	3	2	3

La plupart des autres élèves ont indiqué ce qu'ils pensaient être un ordre de grandeur du nombre de simulations à effectuer, par des locutions telles que « plusieurs », « un grand

nombre » etc.
Nous avons répertorié ces réponses dans le tableau suivant :

formulation	**plusieurs**	**beaucoup, nombreuses**	**un grand nombre**	**un très grand nombre / de très nombreuses fois**
effectif	10	3	9	3

Remarquons que pour deux des trois élèves qui proposent de faire 1000 simulations, l'un dit « un très grand nombre » de simulations, l'autre « de nombreuses simulations ».
Les deux élèves qui proposent de faire 100 simulations considèrent pour l'un que c'est « un grand nombre » de simulations, pour l'autre qu'il s'agit de « plusieurs » simulations.
L'un des trois élèves qui proposent de faire 10 simulations dit faire « plusieurs » simulation.
Enfin, l'un des deux élèves qui proposent de faire plus de 1000 simulations annonce « un très grand nombre » de simulations, l'autre « un grand nombre ».

Le nombre de simulations à faire paraît relativement vague, ce qui semble refléter les limites des programmes des années 2000 quant à la présentation suggérée de la loi des grands nombres.

Signalons que certains élèves se sont inquiétés du fait que la valeur moyenne calculée pourrait ne pas être un nombre entier, et ont su trouver une solution technique (partie entière, arrondi) à ce problème.

5. L'incertitude

Nous citons la réponse à la partie B. qui nous semble la plus aboutie de toute l'expérimentation, celle de l'élève n° 89, qui cependant n'a pas perçu, pendant l'expérimentation, que la présence de trois joueurs permet d'envisager <u>plus</u> de cases sur l'échiquier.
Cette élève a commencé la question A.1. en simulant un jeu avec 3 joueurs : dans les cellules d'adresses A1, B1 et C1 elle avait inscrit les libellés : « Premier joueur », etc, et dans les cellules A2, B2 et C2 la formule :
ABS(ALEA.ENTRE.BORNES(1 ;4) – ALEA.ENTRE.BORNES(1 ; 4)).
On peut penser que le choix d'une formule condensée pour cette élève était une nécessité qui

a pu s'imposer lorsqu'elle a voulu gérer les trois joueurs. Elle a rapidement réalisé qu'elle était hors sujet (c'est dommage, il eût été intéressant de voir la suite…), et nous a demandé une autre feuille pour recommencer.

> *« • On calcule l'espérance pour un tour d'un joueur, cette espérance représente l'espérance du nombre de case qu'il peut parcourir*
> *E(X) = 0×2/8+1×3/8+2×2/8+3×1/8 = 3/8+4/8+3/8 = 10/8 = 1,25 (cases)*
> *• En multipliant cette espérance valable pour un tour, nous trouverons l'espérance de nombre de cases parcourues à la fin des 20 tours :*
> *E(X) × 20 = 1,25 × 20 = 25 (cases)*
> *On peut donc considérer que le nombre de cases « idéales » parcourues au bout de ces 30 mn par un joueur est 25 cases. Cependant pour plus de sécurité il serait préférable que Monsieur Dupont choisisse un nombre plus faible de case. En effet nous ne savons pas à quelle amplitude ce nombre de cases varies globalement (écart-type ?) et l'amplitude maximale des valeurs est importante : 3 × 20 – 0 × 20 = 60. Même si la probabilité de n'avoir que des écarts égaux à 0 ou à 3 est plus faible, celle d'avoir une espérance juste un peu inférieur à 25 sur 20 parties n'est pas négligeable et sur les trois joueurs il est tout à fait probable que les 3 ont une espérance inférieure à 25. Donc selon moi il est préférable d'abaisser le nombre de cases à 20 pour que la probabilité que sur les 3 joueurs au moins l'un atteigne le bout du plateau et par conséquent la fin du jeu soit plus « sûre » c'est à dire plus élevée*
> *Bien sûr le plateau où il est sûr que les joueurs atteignent la fin : celui à zéro case fait perdre au jeu son intérêt, c'est pour cela qu'il vaut mieux éviter de diminuer trop sensiblement le nombre de cases. »*

5. La prise en compte du nombre de joueurs

Cinq élèves ont essayé de prendre en compte le fait qu'il y avait plusieurs joueurs, de façons plus ou moins pertinentes.

Remarquons que d'une part les questions intermédiaires n'aidaient pas à prendre en compte ce paramètre puisque la partie A. du sujet demandait de calculer la loi de probabilité de X, nombre de cases avancées sur un tour par <u>un</u> joueur, et de simuler 20 tours pour <u>un</u> joueur, et que d'autre part la question B. était formulée sans évoquer le nombre de joueurs.

La question est difficile.
Nous l'avons évoquée avec nos deux classes (première et terminale) de l'année scolaire 2009-2010 immédiatement après l'expérimentation. La plupart des élèves ont eu du mal à imaginer que la présence de trois joueurs augmentait le nombre de cases que le meilleur d'entre eux pouvait atteindre en 20 tours.
Aucun élève n'a correctement géré ce paramètre lors de notre expérimentation.

Nous pensons que cette question place dans une situation où précisément les simulations de l'expérience aléatoire complète en jeu (par une imitation au plus près d'une partie avec trois joueurs) permettent de résoudre le problème, là où la modélisation probabiliste est en défaut (que vaut l'espérance de la variable aléatoire Max(X,Y, Z) connaissant les espérances de X, de Y et de Z ?). Ici les issues sont trop nombreuses pour que le calcul soit raisonnable.
Une simulation complète du jeu avec trois joueurs dans un langage impératif n'est pas difficile.

6. En conclusion

Nous souhaitions voir si des élèves, dans une situation d'incertitude liée à une expérience aléatoire qu'ils ont simulée et dont ils ont déterminé la loi de probabilité, utilisent cette simulation ou cette loi de probabilité, et dans l'affirmative, de quelles façons.

L'exercice était d'autant plus difficile qu'il était à faire en temps très limité (50 minutes au plus) et que les élèves pouvaient imaginer des simulations, mais non les réaliser sur ordinateur : la variabilité de la valeur moyenne serait ainsi apparue, amenant peut-être pour certains élèves des questionnement nouveaux, en particulier sur le niveau d'incertitude du choix final.

Nous pouvons faire ressortir de l'étude des réponses à la question B. quelques points :

• un nombre important d'élèves a su réinvestir les outils enseignés dans le cours de statistique-probabilités en dehors des savoir-faire habituellement évalués, en prenant des initiatives qui n'étaient pas suggérés dans l'énoncé : calcul d'une espérance, et même calcul de l'espérance d'une variable aléatoire à partir de la loi de probabilité d'une autre variable aléatoire, calcul d'une valeur moyenne d'une série statistique obtenue par répétition d'une simulation.

• Quelques élèves manifestent une assez bonne compréhension du lien entre l'espérance mathématique d'une variable aléatoire et sa valeur moyenne observée à l'issue de n épreuves via la loi des grands nombres (ce point figure dans les programmes de première S).
Mais certains élèves envisagent ce lien en termes d'égalité, ce qui nous suggère tout à la fois une mauvaise prise en compte de la variabilité de la valeur moyenne observée sur un échantillon, une mauvaise interprétation de l'espérance mathématique d'une variable aléatoire, et peut-être une mauvaise compréhension de la convergence en probabilité qu'évoque la loi des grands nombres.

• Les termes choisis par les élèves pour répondre à la question B.(« comment aider Monsieur Dupond ? ») font souvent apparaître que, puisque la question posée était une aide à la décision **dans une situation d'incertitude, le résultat de la décision restera incertain** (incertain ici quant au fait que la partie se terminera dans le temps imparti) ; l'incertitude apparaît au travers de formulations telles que « environ », des modes conditionnels (« devrait »), ou de la reprise du verbe « espérer » qui se trouvait dans l'énoncé.
Quelques élèves ont essayé de contrôler cette incertitude, mais toujours dans les termes : on va se donner une marge, pour « augmenter la certitude », mais aucun élève n'a tenté de poser le problème dans les termes : « il y aura toujours une incertitude, essayons de mesurer le risque que l'on peut associer aux diverses décisions possibles, pour permettre à Monsieur Dupond de choisir le risque qu'il veut bien assumer ».

Nous allons montrer dans la quatrième partie que ce type de problème que nous avons choisi pour mener notre expérimentation, pourrait être une situation pour introduire, par une étude statistique, la notion de risque.

QUATRIEME PARTIE

Un autre modèle pour une expérience aléatoire

Nous allons nous inscrire dans la théorie des « quatre t » de Chevallard (1999) pour montrer qu'une *simulation informatique* d'une expérience aléatoire donnée « *est une* **machine** *dont la mise en fonctionnement permet* **de produire des connaissances** *relatives au système modélisé* », donc, peut-être, un modèle en ce sens.

Dans le chapitre 19, nous reprendrons le jeu de M. Dupond, et montrerons comment utiliser la simulation informatique de ce jeu avec trois joueurs, pour déterminer un nombre de cases pour lequel on peut estimer que le jeu sera terminé à l'issue de 20 tours, avec un risque d'erreur contrôlé.
Nous procéderons à une analyse des résultats obtenus avec cette simulation informatique (y compris l'évaluation du risque) à la lumière de la théorie classique des probabilités.
Nous terminerons ce chapitre en remarquant que cette simulation ne met en œuvre, en dehors des hypothèses de modélisation décrites dans la partie 2 et la partie 3 de notre thèse, aucun élément de la théorie classique des probabilités, et constitue un modèle alternatif au modèle probabiliste classique.

Dans le chapitre 20, nous tenterons de détailler les organisations praxéologiques associées aux tâches de *modélisation probabiliste* et de *simulation informatique* d'une expérience aléatoire donnée, et de préciser les connaissances que l'un ou l'autre des deux modèles peut apporter à la connaissance de l'expérience de départ.

Pour éclairer la suite de notre propos, nous interrogerons, dans le chapitre 21, les expressions « esprit probabiliste » et « esprit statistique » rencontrées chez certains auteurs.

Nous essayerons alors, dans le chapitre 22, de préciser des objectifs didactiques que l'on peut assigner aux simulations dans le cadre de l'enseignement des probabilités au lycée [69], dans la

[69] nous laisserons de côté les objectifs didactiques possibles dans le cadre de l'enseignement de l'algorithmique-

perspective de proposer aux élèves des situations problèmes leur permettant de s'exercer à une démarche probabiliste.

Enfin dans le chapitre 23, nous montrerons que les simulations informatiques d'expériences aléatoires [70] ne constituent pas « un cercle vicieux didactique », si l'on veut bien ne pas s'enfermer dans l'idée que, puisqu'elle n'est simulation que si elle a même modèle probabiliste que l'expérience d'origine, une simulation devrait être construite à partir de ce modèle, ou autour de ce modèle : la possibilité de faire de la tâche de simulation informatique dévolue aux élèves, une activité de type algorithmique, et non plus une seule activité de choix de paramètres dans un tirage d'urne, permet de poser dans d'autres termes la preuve de l'équivalence entre l'expérience d'origine et sa simulation informatique « imitation au plus près ».

---

programmation, même s'ils nous paraissent loin d'être négligeables, et s'il nous paraît difficile de dissocier les deux cadres.

[70] Nous nous sommes limitée à celles qu'il nous paraît raisonnable d'envisager en lycée.

CHAPITRE 19

Une solution au problème de M. Dupond basée sur une étude statistique

1. Etude statistique des résultats observés en répétant la simulation d'une partie avec trois joueurs

Il s'agit donc, dans ce problème, de choisir un nombre de cases de telle sorte qu'une partie avec trois joueurs se termine en au plus 20 tours, en fait implicitement : en environ 20 tours.

• Si l'on veut au plus 20 tours, on peut y arriver « presque certainement » avec une ou deux cases, mais le jeu n'aura aucun intérêt.

• Si l'on veut environ 20 tours, cela fixe une durée (ce qui est souvent indiqué sur les jeux de société, par la mention d'une « durée moyenne »).

Dans cette expérimentation, les élèves qui ont répondu à la question B (« comment aider M. Dupond ») ont presque tous eu l'idée de s'appuyer sur des résultats obtenus soit en modélisant, soit en simulant 20 lancers pour **un** joueur ; cinq élèves ont tenté de prendre en compte le fait que l'énoncé mentionnait trois joueurs.

Appelons X le nombre de cases avancées par **un** joueur ; l'espérance est $E(X) = 25$, et si on répète un grand nombre de fois la simulation, le nombre moyen de cases avancées est voisin de 25.

Qu'est-ce qui change avec trois joueurs ?

Appelons X_1, X_2, et X_3 le nombre de cases avancées par chacun des trois joueurs. $E(X_i) = 25$.

Appelons $Y = \max \{X_1, X_2, X_3\}$. Au bout de 20 tours, l'un des joueurs a parcouru Y cases, il a été plus loin que les autres.

C'est sur Y qu'il faut travailler si l'on veut que la partie dure environ 20 tours.

Que sait-on de Y ?

$Y \geq X_i$, donc $E(Y) \geq E(X_i) = 25$.

Peut-on calculer E(Y) ?

A notre connaissance, il n'y a pas de façon simple de calculer l'espérance de la variable aléatoire définie comme max de plusieurs variables aléatoires dont on connaît l'espérance.

Dans le contexte qui nous intéresse, le calcul est théoriquement faisable : il relève d'une

combinatoire simple, mais disons que l'arbre de probabilités commence à devenir forêt.
Nous n'avons pas tenté ce calcul de l'espérance de Y, d'autant que les simulations nous en donnent une valeur approchée comme nous le justifierons au paragraphe 2.

Aucune des solutions proposées par les élèves ne permet de choisir un nombre de cases avec la certitude que le jeu se terminera bien à l'issue de 20 tours.
Mais peut-on choisir un nombre de cases x avec un risque inférieur à 5 % (par exemple) que la partie avec 3 joueurs ne soit pas terminée à l'issue des 20 tours ?
Autrement dit, peut-on choisir un nombre de cases x avec un risque de 5 % que la partie ne soit pas achevée en au plus 20 tours ?

Oui, avec une étude statistique d'un grand nombre de simulations.
Evidemment, il faudra, à un moment approprié, *définir mathématiquement* ou *justifier* par un discours d'ordre technologique cette notion de *risque* et sa quantification.

Plus précisément, pour le problème de M. Dupond :
• On simule 20 tours avec trois joueurs : chacun avance de X_1, X_2 et X_3 cases à l'issue de ces 20 tours.
$Y = \max(X_1, X_2, X_3)$
Ceci peut se programmer sur <u>une</u> ligne du tableur [71] :
• On répète 1000 (ou 3000) fois
En recopiant la première ligne vers le bas
• On peut faire alors une étude statistique des valeurs prises par Y,
par exemple on peut calculer le 5$^{\text{ème}}$ centile C5 (c'est à dire la plus petite valeur de la série telle qu'au moins 5 % de l'effectif soit avant cette valeur ; alors au moins 95% de l'effectif est au delà puisqu'il s'agit d'inégalités au sens large).
Ceci se fait facilement avec un tableur
La mise en œuvre de cette méthode nous a donné
• une valeur moyenne pour Y autour de 28,6 (et une valeur moyenne pour les X_i voisine de

[71] On peut aussi établir une feuille de calcul pour une partie de 20 tours avec 3 joueurs (une ligne par tour par exemple, comme dans notre expérimentation), puis utilisant le caractère « langage orienté objet » de VBA, programmer le re-calcul de cette feuille un très grand nombre de fois, les résultats étant stockés dans une feuille de calcul pour permettre ensuite un traitement statistique.

25, valeur attendue)
• C5 à 23 ou 24 (il y a des fluctuations, bien sûr, lorsqu'on lance des « re-calculs » de la feuille).

2. Prise de décision sur la base d'un risque calculé

Les observations sur ce grand nombre de *simulations au plus près* sont représentatives de « la réalité » de cette expérience aléatoire qu'est la simulation. Les deux variables Y et Y' égales au nombre de cases avancées par celui des trois joueurs qui va le plus loin, dans le jeu et dans sa simulation, ont la même loi de probabilité, puisque les deux expériences (le jeu et sa simulation) ont même modèle (ceci est encore un postulat). Les observations sur Y' (variable associée à la simulation) vont pouvoir remplacer les observations sur Y (variable associée au jeu).
Donc avec k = C5 cases, il y a 5% des parties au plus, dans le jeu et dans la simulation, qui ne se terminent pas dans le temps imparti, puisque pour 95% des parties au moins, Y, le nombre de cases avancées par celui des trois joueurs qui a été le plus loin, a dépassé C5.

On peut *définir* ce pourcentage comme *le risque* que la partie se termine hors délai.

Monsieur Dupond peut alors prendre, avec un risque $\alpha = 5\%$ que la partie ne soit pas terminée dans les délais, la décision de faire son jeu avec k_5 = C5 cases.

Monsieur Dupond peut décider de prendre moins de risque, et préférer évaluer ce risque par $\alpha = 1\ \%$. Il prendra alors la décision de faire son jeu avec k_1 = C1 cases.

Evidemment, C1 $\leq$ C5, et les prises de décisions dans les termes ici présentés sont cohérentes : moins on prend de cases sur le jeu, moins il y a de risque que la partie ne s'achève pas dans les délais impartis.

C'est en ce sens que nous pouvons parler d'une prise de décision sur la base d'un *risque calculé* : M. Dupond peut choisir une plus ou moins grande prise de risque, le risque nul consistant, comme l'a remarqué l'élève n°89, à placer k = 0 case sur le plateau.

Remarquons d'abord que la simulation d'une partie de 20 tours avec trois joueurs a permis,

par sa répétition un grand nombre de fois, de produire des connaissances relatives à l'expérience aléatoire qui consiste à faire avancer pendant 20 tours trois joueurs sur le plateau, et à regarder de combien de cases a avancé celui des trois joueurs qui va le plus loin.
On peut alors dire que cette simulation est un modèle de l'expérience initiale puisqu'elle permet de produire des connaissances sur cette expérience (cf chapitre 1, paragraphe 3).

Il importe que la simulation soit informatique, pour permettre raisonnablement sa répétition un grand nombre de fois.

Remarquons ensuite, que dans ce paragraphe, nous nous en sommes tenue à la simulation informatique de la répétition un grand nombre de fois de la simulation d'une partie de 20 tours avec 3 joueurs, et à une étude statistique des résultats observés, évitant toute allusion à une probabilité ou à la loi des grands nombres, autrement dit évitant toute référence à la théorie mathématique des probabilités.

Cela ne veut pas dire, bien entendu, que cette théorie n'a rien à dire sur ce qui vient d'être décrit.

3. Interprétation probabiliste de ce risque α

3.1. Une interprétation « faible »

Appelons $C\alpha$ le centile de rang α, et considérons un jeu comportant $k = C\alpha$ cases, et répétons n fois une partie de 20 tours avec trois joueurs. Appelons p_k la probabilité qu'une partie soit terminée à l'issue des 20 tours (c'est à dire que l'un des joueurs au moins ait parcouru les k cases), et q_k la probabilité de l'événement contraire.
Définissons la variable aléatoire X_n égale au nombre de parties achevées à l'issue des 20 tours.

La loi des grands nombres nous dit que $F_n = \frac{X_n}{n}$, qui est la fréquence observée sur n parties de l'événement « la partie est terminée à l'issue des 20 tours », *tend en probabilité* vers p_k.
Pour $\alpha = 0{,}05$, sur n = 3000 parties simulées, nous avons observé une fréquence F_n de parties achevée à l'issue des 20 tours d'au moins 0,95 en prenant k cases, avec k = C5 = 23.
Comme F_n converge en probabilité vers p_k, nous pouvons conclure que, le nombre de parties simulées étant considéré comme suffisamment grand, p_k est supérieur ou égal à 0,95 , et par

conséquent que la probabilité q_k que la partie ne soit pas terminée à l'issue des 20 tours est inférieure à $\alpha = 0,05$.

L'indicateur α défini comme indicateur de risque dans le paragraphe précédent peut être interprété en terme de probabilité comme nous allons le montrer.

Remarquons que les considérations qui précèdent peuvent être tenues dans le cadre d' un énoncé de la loi des grands nombres considéré comme un théorème de la théorie des probabilités, et donc lié au *modèle probabiliste* choisi [72], ou comme une *loi empirique extérieure* à la théorie des probabilités.

Dans la mesure où cette interprétation peut se faire à l'aide de la seule notion de loi de probabilité sur les deux issues : « la partie est terminée à l'issue des 20 tours » et « la partie n'est pas terminée », et d'une loi qui peut être présentée comme extérieure à la théorie des probabilités, nous avons parlé d'« interprétation faible »

3.2. Une interprétation plus forte

Reprenons l'étude de X_n : X_n suit une loi binomiale de paramètres n et p_k , qui converge lorsque n tend vers l'infini vers une loi normale (si p_k et $q_k = 1 - p_k$ sont du même ordre de grandeur) d'espérance n p_k et de variance n p_k q_k . Dans la pratique, on remplace la loi binomiale par la loi normale pour $n > 30$, $n\, p_k > 5$ et $n\, q_k > 5$, par conséquent sous les mêmes hypothèses, la variable $F_n = \frac{X_n}{n}$ suit une loi normale d'espérance $\mu = p_k$ et de variance $\sigma^2 = p_k\, q_k$.

La variable centrée réduite $Z = \frac{F_n - \mu}{\sigma}$ suit la loi normale $\mathcal{N}(0 ; 1)$.

Pour ß dans]0 ; 1[, il existe un unique réel positif $t_ß$ tel que $P(- t_ß \leq Z \leq t_ß) = 1 – ß$. [73]

---

[72] Ici le choix se fait sur les *hypothèses de modélisation* sur les deux dés utilisés pour jouer, à savoir qu'ils sont bien équilibrés, et il n'y a pas d'autre hypothèse possible lorsqu'on en est à concevoir le jeu.

Ce serait différent si la question était posée en rapport avec un jeu particulier, commercialisé par M. Dupont, contenant deux dés, et pour lequel on voudrait étudier le risque qu'une partie ne soit pas terminée à l'issue des 20 tours. Une étude d'expériences liés à ces deux dés particuliers pourrait amener une hypothèse de modélisation autre que la loi équirépartie pour chacun des dés.

[73] Ce théorème figurera peut-être dans les programmes de Terminale S des années 2010, cf les projets de programme en date du 8 mars 2011, sur le site « Eduscol »

Pour $ß = 0,05$, $t_ß \sim 1,96$, et pour $ß = 0,01$, $t_ß \sim 2,58$.

On en déduit [74] que la limite quand n tend vers l'infini de la probabilité P_n que $F_n = \frac{X_n}{n}$ soit dans l'intervalle I_n est de $1 - ß$,

où I_n est l'intervalle $[p_k - t_ß \frac{\sqrt{p_k (1 - p_k)}}{\sqrt{n}} ; p_k + t_ß \frac{\sqrt{p_k (1 - p_k)}}{\sqrt{n}}]$

Fixons dans ce qui suit $ß = 0,05$; pour n assez grand, $\frac{F_n - \mu}{\sigma} = Z$ appartient à l'intervalle $[- 1,96 ; 1,96]$ avec une probabilité de 0,95, et donc, avec une probabilité égale à 0,95 ,

$\mu = p_k$ est dans l'intervalle $[F_n - 1,96 \frac{\sqrt{p_k (1 - p_k)}}{\sqrt{n}} ; F_n + 1,96 \frac{\sqrt{p_k (1 - p_k)}}{\sqrt{n}}]$.

Une étude de la fonction f définie sur $[0 ; 1]$ par $f(x) = x (1-x)$ montre que f(x) est majoré sur $[0 ; 1]$ par 1/2.

Par conséquent $0 < 1,96 \frac{\sqrt{p_k (1 - p_k)}}{\sqrt{n}} < 1$, et donc $I_n \subset J_n = [F_n - \frac{1}{\sqrt{n}} ; F_n + \frac{1}{\sqrt{n}}]$.

Donc par exemple en prenant n supérieur à 2500, la probabilité que F_n approche p_k avec une précision d'au moins $1/50 = 2.\ 10^{-2}$ est égale à 0,95.
La probabilité que l'écart entre F_n et p_k dépasse $2.\ 10^{-2}$ est de 0,05.

La feuille de calcul composée pour faire l'étude statistique nous permet de déterminer F_n pour n = 3000, et donne F_{3000} supérieur à 0,95.[75]

Plus précisément : avec n = 3000, C5 est encore soumis à la fluctuation d'échantillonnage [76]
Nous avons trouvé (résultats arrondis à deux ou trois décimales, au plus proche) :

---

[74] idem

[75] Ce qui était prévisible, k ayant été déterminé comme le 5ème centile, au moins 95 % des données sont au dessus de C5 (ou égales), donc au moins pour 95 % des parties, au moins un joueur a franchi C5 cases ou plus, donc dans au moins 95 % des parties, la partie est achevée à l'issue des 20 tours.

[76] C5 est un entier. Sa stabilisation lorsqu'on répète un grand nombre de fois l'expérience finira par masquer la fluctuation d'échantillonnage, qui continuera de s'observer pour F_n, tant que l'outil de calcul permettra cette observation.

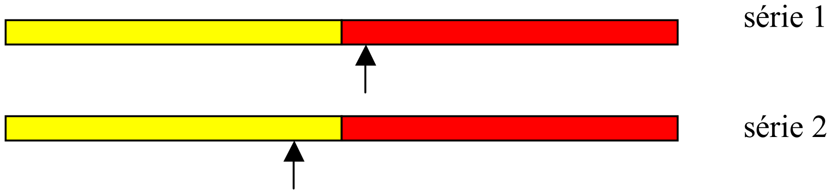
3000 parties / indicateurs	essai 1	essai 2	essai 3
E(Y)	28,64	28,73	28,64
k = C5	24	23	24
F_n	0,953	0,980	0,953
J_n	[0,933 ; 0,973]	[0,960 ; 1]	[0,933 ; 0,973]

Selon la valeur de F_n retenue, on conclura que p_k dépasse 0,93 ou que p_k dépasse 0,96, avec une probabilité de 0,95.

Les hypothèses pour considérer que X_n suit une loi normale sont satisfaites : 3000 > 30 et 3000 × 0,95 > 5 et 3000 × 0,05 > 5.

Les intervalles J_n auraient pu être affinés (majoration de 1,96 /2 par 1 et majoration du produit $p_k\ q_k$ par 0,5), mais cette amélioration est illusoire : C5 comme F_n sont soumis à la fluctuation d'échantillonnage.

La plus forte valeur obtenue pour p_k avec k = 23 s'explique par une sorte d'effet de bord : le plus souvent C5 est 24 ; 23 s'obtient plus exceptionnellement. On peut représenter différentes séries statistiques :



Le rectangle jaune représente les valeurs de la série des Y observés à 23, et le rectangle rouge les valeurs à 24. On peut penser que le « curseur 5% de l'effectif », représenté par la flèche noire, est toujours au voisinage de la frontière 23/24.

Pour la série 1, on aura C5 = 24, et un peu plus de 95 % de l'effectif de la série sera supérieur ou égal à 24, tandis que pour la série 2, on aura C3 = 23, et un pourcentage nettement supérieur à 95 % sera au delà de 23.

On peut cependant conclure, malgré la variabilité des résultats observés, que la probabilité ß est forte (0,95) qu'avec un nombre de cases k = 24, la probabilité qu'une partie s'achève à l'issue des 20 tours soit voisine de 0,95, et que donc la probabilité est forte (0,95) que la probabilité que la partie ne se termine pas soit de l'ordre de cette valeur α = 0,05 déterminée statistiquement.

Avec n plus grand, la fluctuation de C5 n'est pas apparue (plusieurs essais) ; nous reproduisons les résultats observés dans les trois premiers :

10000 parties / indicateurs	essai 1	essai 2	essai 3
E(Y)	28,70	28,64	28,67
k = C5	24	24	24
F_n	0,951	0,950	0,950
J_n	[0,941 ; 0,961]	[0,940 ; 0,960]	[0,940 ; 0,960]

Ces résultats confortent la conclusion : si l'on met k = 24 cases sur l'échiquier, avec une probabilité ß = 0,95, la probabilité que la partie se termine dans les temps impartis, c'est à dire 20 tours, est égale à 0,95 à 10^{-2} près.
Autrement dit, il est très probable (probabilité ß = 0,95) que le risque α (estimé statistiquement au paragraphe 1) que la partie ne soit pas achevée dans les temps avec k = 24 cases soit égal à 0,05.

4. La simulation a permis de produire des connaissances

Nous venons de voir que la simulation d'une partie de 20 tours avec trois joueurs a permis de produire des information nouvelles sur le jeu, permettant des prises de décision.
Nous allons étudier la nature de ces informations.
Rappelons d'abord que la simulation n'est réellement exploitable que parce qu'il s'agit d'une simulation informatique, qui peut être répétée un grand nombre de fois, avec conservation des résultats [77], ce qui permet une étude statistique sur les résultats empiriques recueillis.
Les échantillons étaient de taille n = 2500 ou 10 000.
Cette *étude statistique* a pu porter

- sur la valeur moyenne de la série statistique des n valeurs de la variable aléatoire Y égale au nombre de cases parcourue par le joueur qui a été le plus loin à l'issue des 20 tours.
- sur les centiles de cette série

[77] Ceci nécessite donc un outil informatique possédant suffisamment de mémoire, ce qui peut ne pas être le cas avec une calculatrice de lycée.

- la valeur de k, nombre de cases sur l'échiquier, étant fixée, on a pu calculer la fréquence des parties achevées à l'issue des 20 tours, et définir un *indicateur de risque* que la partie ne soit pas achevée à l'issue des 20 tours.

L'exploitation de cette simulation informatique a fait ressortir qu'il demeure des résultats incertains, liés à la fluctuation d'échantillonnage.

L'interprétation des résultats obtenus peut se faire en dehors de la théorie des probabilités, avec une notion « faible » de probabilité en termes de *chances*, ou avec la *loi (faible) des grands nombres* [78] .

L'étude statistique peut aussi être replacée dans le cadre de la théorie des probabilités, avec l'éclairage que nous avons donné dans le paragraphe précédent.

La situation un petit peu particulière que nous venons d'étudier [79] ne doit pas faire oublier cependant que la simulation, comme la modélisation passe toujours par des *hypothèses de modélisation* , dont il faut rendre compte, hypothèses qui sont susceptibles d'être remises en cause, par exemple par confrontation avec le système réel étudié quand c'est possible. Cette confrontation, si l'on s'en tient à une simulation informatique du type *imitation au plus près*, sera une confrontation des données empiriques obtenues par la répétition de l'expérience avec des données expérimentales obtenues à partir de l'expérience aléatoire elle-même, s'il y en a.

Cette confrontation peut se faire soit en considérant les résultats expérimentaux de l'expérience initiale comme ceux de la simulation, et en comparant les indicateurs statistiques des deux séries, et on reste alors en dehors de la théorie des probabilités, soit en comparant les résultats avec des tests de la théorie des probabilités (test du χ^2 ou autre).

L'étude de cet exemple du jeu de M. Dupond nous amène à poser l'hypothèse que la *simulation informatique* d'une expérience aléatoire est un modèle du système réel que

---

[78] *loi*, extérieure à la théorie des probabilités : « la fréquence tend à se stabiliser autour d'un nombre qu'on peut appeler la probabilité empirique », par opposition au *théorème des grands nombres*, qui est un théorème de la théorie des probabilités.

[79] cf.χ ce que nous avons expliqué concernant les hypothèses de modélisation dans le paragraphe précédent.

constitue cette expérience aléatoire, au même titre que le *modèle probabiliste* de cette expérience.

CHAPITRE 20

Praxéologies

Nous venons de postuler que la *simulation informatique* d'une expérience aléatoire peut être aussi considérée comme un *modèle* de cette expérience dans la mesure où elle permet *aussi* de produire des connaissances sur l'expérience aléatoire ainsi modélisée (chapitre 1, paragraphe 3).
Nous allons tenter de voir les liens qu'entretiennent (ou que n'entretiennent pas) ces deux types de modèles, au delà du simple constat de la partie 2 de cette thèse que la détermination de ces deux modèles (*probabiliste* et *informatique*) passe par une étape commune que nous avons appelée *pré-modélisation* comportant la détermination des *hypothèses de modélisation*,.

Nous allons dans ce chapitre considérer une expérience aléatoire que nous noterons E.A., et étudier les deux tâches t_m : modéliser la tâche E.A. au sens du *modèle probabiliste*, et t_s : simuler la tâche E.A.
t_m relève du type de tâches T_m : *modéliser une expérience aléatoire* e.a.
t_s relève du type de tâches T_s : *simuler une expérience aléatoire* e.a.

Nous allons nous placer dans la théorie des « 4t » de Chevallard, et tenter de préciser les techniques, technologies et théories qu'on pourrait associer à ces deux types de tâche.

Nous présenterons d'abord, dans des tableaux, des *praxéologies ponctuelles* pour ces deux types de tâches, en les commentant. Ces praxéologies se rapportent, pour la tâche de *modélisation probabiliste*, à l'institution classe I, selon les « programmes 2000 » : I_P pour la classe de première S, I_T pour la classe de terminale S.

1. Praxéologies pour T_m

1.1. Expériences aléatoires simples

EXPERIENCES ALEATOIRES SIMPLES			
I	**techniques τ**	**technologies θ**	**théories Θ**
I_P I_T	• **τ_{m1} : approche Laplacienne** - choix des issues $x_1, \ldots, x_n$ - choix des hypothèses de modélisation sur ces issues : pour des raisons de symétrie les issues sont équiréparties - dénombrement des issues favorables et possibles (arbres, tableaux, utilisation de *théorèmes de dénombrement*) - pour une loi équirépartie, $p(\{x_i\}) = \frac{\text{nbre cas favorables à } x_i}{\text{nbre cas possibles}}$ et $p(A) = \frac{\text{nbre cas favorables à A}}{\text{nbre cas possibles}}$ • **τ'_{m1} : *approche avec expertise*** - choix des issues $x_1, \ldots, x_n$ - choix des hypothèses de modélisation sur ces issues, certaines probabilités étant fournies par l'extérieur [80]	• θ_{m1} : - Un univers fini Ω qui décrit les issues possibles. $\Omega = \{x_1, \ldots, x_k\}$ - un événement A est une partie de Ω auquel on peut associer sa probabilité p(A) $\in [0\,;1]$ - p est complètement déterminée par les $p(x_i)$ - p est additive - une loi est équirépartie si toutes les issues ont même probabilité : 1/k. - éventuellement loi obtenue à partir d'une autre loi (souvent équirépartie) par « transport par une variable aléatoire » [81]	• Θ_{m1} : - espace probabilisé fini $(\Omega\,;\mathcal{P}(\Omega)\,;p)$; espérance et écart type d'une loi ; variable aléatoire, espérance mathématique et écart type. - analyse combinatoire : p-listes, permutations, combinaisons

[80] On fait appel à des *experts* pour estimer certaines probabilités.
Lorsque la loi n'est pas équirépartie ou ne peut pas se déterminer à partir d'une loi équirépartie, les exercices *classiques* renvoient implicitement à cette expertise extérieure, sans donner d'indication sur la manière dont sont alors estimées les probabilités : par exemple « on estime que la probabilité pour que le test soit positif sachant que la personne est malade est 0,90 » ou « la face 6 a deux fois plus de chances de sortir que les autres faces ».

[81] cf. les programmes 2000. C'est le cas par exemple pour la somme de deux dés.

1.2. Expériences aléatoires composées

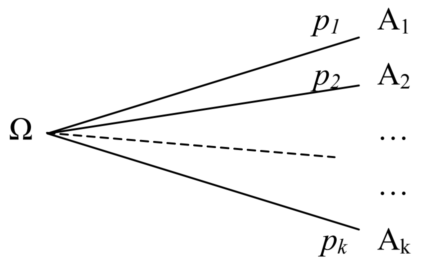
EXPERIENCES ALEATOIRES COMPOSEES			
I	**techniques τ**	**technologies θ**	**théories Θ**
I_T	• **τ_{m2} : arbre de probabilité** C'est une technique utilisée pour des expériences composées. On construit de manière récursive l'arbre de probabilité en décomposant l'univers (événement certain) en une partition d'événements A_i (i.e. événements deux à deux incompatibles dont la réunion est Ω) dont on détermine les probabilités (en faisant des hypothèses de modélisation), puis en recommençant pour chacun de ces événements A_i considérés comme nouvel univers (ce qui mettra en jeu des probabilités conditionnelles : $p_{A_i}(B_j)$, etc.). La loi de probabilité sur les issues intéressantes est obtenue à partir de l'arbre de probabilité, en mettant en œuvre les règles de calcul dans un arbre de probabilité (voir ci-dessous).	• θ_{m2} : Un arbre de probabilité est un arbre dont les nœuds sont étiquetés par des événements, et les branches par des probabilités. les arbres de probabilité sont définis de façon récursive (cf. plus bas). La racine est l'univers, c'est à dire l'événement certain. Il faut, pour étiqueter les branches, faire des *hypothèses de modélisation* en décidant si des événements sont indépendants ou non (indépendance *de bon sens* [82]), et en évaluant des probabilités conditionnelles, par une approche laplacienne, ou en faisant appel à une expertise extérieure .	• Θ_{m2} : espace probabilisé $(\Omega ; \mathcal{P}(\Omega) ; p)$ + les probabilités conditionnelles, et la formule des probabilités totales, qui justifie les règles de calcul dans un arbre de probabilité. NB : la notion d'indépendance stochastique n'est pas nécessaire.

[82] non calculée comme indépendance stochastique.

1.2.1. La définition récursive d'un arbre de probabilité

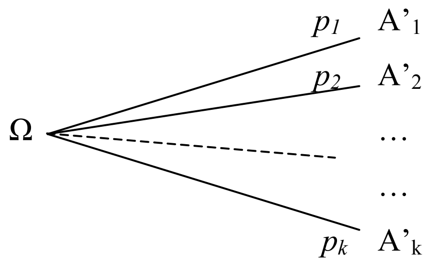
Un arbre de probabilité est

• un arbre tel que :



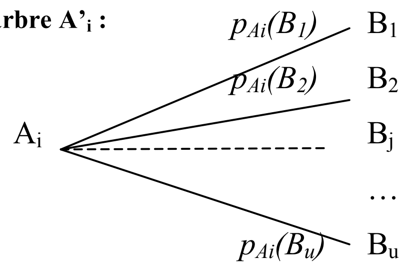
où les A_i forment une partition de l'univers Ω, et $\sum_i p_i = 1$ (c'est à dire que la somme des étiquettes des branches partant de la racine est égale à 1),

• ou un arbre tel que :



où les A'_1, A'_2, …, A'_k sont des arbres de probabilité tels que l'arbre A'_i ci-dessous :

arbre A'_i :



c'est à dire que pour cet arbre A'_i, l'univers est A_i, il y a *u* branches étiquetées par les sous-ensembles B_1, …, B_u **de Ω** et $p_{Ai}(B_1)$ représente la probabilité de B_1 sachant A_i ; on a bien entendu $\sum_j p_{Ai}(B_j) = 1$.

L'événement étiqueté à un nœud (respectivement à une feuille) est en réalité l'intersection des événements se trouvant sur les nœuds situés entre la racine et ce nœud (respectivement cette feuille). Ainsi dans le schéma ci-dessus, l'étiquette B_2 correspond à l'événement $A_i \cap B_2$. Les feuilles de cet arbre A'_i sont donc bien des sous ensemble de l'univers A_i.

1.2.2. Les règles de calcul dans un arbre de probabilité

- la probabilité de l'événement E est la somme des probabilités des événements associés aux feuilles de l'arbre, qui sont favorables à E (c'est à dire qui sont inclus dans E) ;
- La probabilité d'un événement associé à une feuille est égale au produit des probabilités se trouvant sur le chemin entre la racine de l'arbre et cette feuille.

C'est la formule des probabilités totales qui justifie ces règles :

Les feuilles de l'arbre correspondent à événements dont on va pouvoir calculer la probabilité. Ces événements forment une partition de l'univers initial Ω (ça se démontre par récursion sur la profondeur de l'arbre). Appelons a_1 , a_2 , … a_t les événements associés aux feuilles de l'arbre, qui sont, pour chacun d'eux, une intersection (conjonction) d'événements ; or $p(A_i \cap B_j) = p_{A_i}(B_j)\ p(A_i)$; on calcule ainsi de façon récursive $p(a_n)$: la probabilité de l'événement associé à une feuille est le produit des probabilités qui se trouvent sur les branches qui sépare cette feuille de la racine,

et, puisque $p(A) = \sum p(A \cap a_j) = \sum p_{a_j}(A)\ p(a_j)$, si l'événement a_j n'est pas favorable à A, $A \cap a_j$ est vide, de probabilité nulle ; si a_j est favorable à A, $A \cap a_j = a_j$ etdonc $p(A \cap a_j) = p(a_j)$

Finalement, $p(A) = \sum_m p(a_m)$ pour tous les a_m favorables à A.

1.3. Expériences aléatoires simples ou composées

EXPERIENCES ALEATOIRES SIMPLES OU COMPOSEES			
I_T et I_P	• **τ_{m3} : approche fréquentiste** On répète n fois l'expérience (n grand), et on calcule la fréquence $f_{n,i}$ de l'issue x_i. La loi de probabilité est approchée par $(f_{n,1}, \ldots, f_{n,k})$	• θ_{m3} : - $f_{n,i}$ est voisin de la probabilité p_i de $\{x_i\}$ selon la loi des grands nombres ; - si X est une variable aléatoire sur l'univers Ω, la valeur moyenne des valeurs de X observées lors des n essais est voisine de l'espérance mathématique de X ; - de même pour l'écart type.	• Θ_{m3} : espace probabilisé $(\Omega ; \mathcal{P}(\Omega) ; p)$ et loi des grands nombres (loi empirique, ou théorème de la théorie des probabilités) ; la convergence des fréquences est une convergence en probabilité.

Lorsque la loi des grands nombres est considérée comme un théorème de la théorie, la convergence des fréquences empiriques est une convergence à l'intérieur du modèle choisi lors du choix des hypothèses de modélisation.

Dans les programmes en cours jusqu'en 2000, la démarche prescrite pour introduire la notion de probabilité était une approche fréquentielle : il était recommandé de s'appuyer sur un travail expérimental, mené en général en première, faisant apparaître *la relative stabilisation* de la fréquence d'un événement pour introduire la notion de probabilité (cf. la première partie de notre thèse).

Ce travail, s'il était mené, était rapidement laissé de côté pour faire place à des exercices relevant des techniques τ_{m1} et τ'_{m2} (c'est à dire la technique τ_{m2} où la formule des probabilités totales est présentée non sous la forme d'un arbre de probabilités, mais sous sa forme linéaire $p(A) = \sum_i p(A\cap B_i) = \sum p_{B_i}(A)\ p(B_i)$, où les B_i forment une partition de l'univers).

Les simulations préconisées dans les programmes des années 2000 avaient, entre autres objectifs, de rendre concrète cette approche fréquentiste, la répétition de sa simulation informatique remplaçant la répétition de l'expérience aléatoire elle-même.

2. Praxéologies pour T_S

2.1. Simulation du modèle probabiliste

I	**techniques τ**	**technologies θ**	**théories Θ**
	EXPERIENCES ALEATOIRES SIMPLES OU COMPOSEES		
I_P et I_P I_T	• **τ_{s1} : simulation du modèle probabiliste** Une fois établie la loi de probabilité de e.a. : distribution $(p_1, \ldots, p_k)$ associée à l'univers $\Omega = \{x_1, \ldots, x_k\}$, on simule un tirage d'un jeton d'une urne contenant N jetons : n_1 jetons numérotés 1, n_2 jetons numérotés 2, …, n_k jetons numérotés k, de telle sorte que $p_i = \frac{n_i}{N}$.	• θ_{s1} : c'est la définition de la simulation d'une expérience aléatoire donnée à l'institution I par les programmes. La programmation de la simulation d'un tirage d'urne se résume à un tirage d'un entier aléatoire de 1 à N selon la loi équirépartie, puis k tests pour indiquer quelle est l'issue réalisée.	• Θ_{m2} : espace probabilisé fini $(\Omega ; \mathcal{P}(\Omega) ; p)$; espérance et écart type d'une loi ; variable aléatoire, espérance mathématique et écart type.

L'expérience aléatoire ainsi obtenue est **par définition** équivalente à l'expérience aléatoire initiale. Elle en est donc une **simulation**.
Cette technique nécessitant la connaissance de la loi de probabilité de l'expérience initiale n'est pas possible lorsqu'on ne sait pas déterminer cette loi de probabilité.

Elle peut présenter un intérêt didactique pour illustrer la variabilité des résultats observés, en permettant de souligner que, une fois choisies les hypothèses de modélisation, le calcul des probabilités est certes déterministe, mais que le résultat de l'expérience reste incertain lors de ses réalisations, en fait lors de ses simulations.

2.2. Imitation au plus près de l'expérience aléatoire

I	techniques τ	technologies θ	théories Θ
	EXPERIENCES ALEATOIRES SIMPLES OU COMPOSEES		
I_P et I_P	• τ_{s2} **: *imitation au plus près*** - établir un *cahier des charges* décrivant les différentes actions constituant l'expérience aléatoire, et leur enchaînement chronologique, en repérant les expériences aléatoires élémentaires constitutives, et en faisant sur elles des *hypothèses de modélisation*, et en décidant quels sont les résultats liés à l'expérience aléatoire que constitue cette simulation, sur lesquels on veut faire une étude statistique [83] ; - faire une *description d'algorithme* respectant ce cahier des charges, description d'algorithme adaptée au langage de programmation choisi ; - programmer cette *imitation.*	• θ_{s2} : - - Détermination *des lois de probabilité* des expériences aléatoires **élémentaires** constituant l'expérience globale pour pouvoir les simuler avec les nombres pseudo aléatoires. - cette description peut faire appel à des techniques et technologies spécifiques à l'algorithmique (par exemple décrire une itération à partir d'un *invariant de boucle*).	• Θ_s : L'algorithmique [84]

[83] Par exemple, pour le lancer de deux dés : calcul de la somme des points amenés, ou calcul de l'écart des points entre les deux dés, etc.

[84] Nous utilisons ici ce terme pour englober des résultats théoriques aussi divers que la génération de nombres pseudo aléatoires, les preuves de programme ou les rudiments de la programmation, impérative (affectation, test, itération) ou avec tableur (adresses absolues ou relatives, recopies), rudiments qui selon le niveau d'explication concernant le fonctionnement du programme, relèvent de la technologie ou de la théorie.

L'appel à la théorie des probabilités concerne les hypothèses de modélisation, et semble ici assez réduit : les expériences aléatoires élémentaires en jeu sont le plus souvent des paradigmes : lancer un dé équilibré, c'est tirer un entier entre 1 et 6 selon une loi équirépartie ; lancer une pièce, c'est tirer un entier, 0 ou 1, selon une loi équirépartie, etc.
Ou alors, les lois à associer à ces expériences aléatoires élémentaires sont données par une expertise, extérieure (une pièce fabriquée par la machine A a une probabilité de 0,05 d'être défectueuse), ou non : on peut imaginer de simuler le lancer de deux dés dont l'un est pipé ; on peut obtenir une approximation de la loi de probabilité du dé pipé par la fréquence observée de sortie de chacune des faces, en lançant le dé un très grand nombre de fois ; si on considère que cette manipulation fait partie de la tâche t_s, la justification ressort de la loi des grands nombres.

3. L'utilisation du modèle probabiliste

A partir de la loi de probabilité de l'expérience aléatoire, on peut définir les espérances et écart-types de variables aléatoires liées à cette expérience. Ces calculs permettent alors des décisions raisonnées.
Ainsi dans le cas de M.Dupond, la connaissance de l'espérance de la variable Y (nombre de cases avancées par celui des trois joueurs qui est allé le plus loin à l'issue des 20 tours), et de son écart-type <u>aurait</u> permis de choisir le nombre de cases à mettre sur l'échiquier avec un niveau de confiance donné.
Si Y est une variable gaussienne [85], lors de n répétitions de l'expérience, c'est à dire lors de la répétition de n parties de 20 tours avec 3 joueurs, 95 % des valeurs de la série des n valeurs observées pour Y sont dans l'intervalle $[E(Y) - 2\,\sigma\,;\, E(Y) + 2\,\sigma]$.
Si on prend un nombre de cases sur l'échiquier égal à $k = E(Y)$ cases, plus de 95 % des parties se termineront dans les délais impartis : si l'on fait l'hypothèse que l'ensemble des valeurs de Y est symétrique par rapport à E(Y), ce seront 97,5 % des parties (95 % plus la moitié des 5 % correspondant aux valeurs au dessus de $E(Y) + 2\,\sigma$).
Bien entendu ce <u>calcul virtuel</u> (nous n'avons <u>calculé</u> ni E(Y) ni σ) n'est ici que pour illustrer ce que l'on aurait pu faire dans le problème posé par M. Dupond avec le seul modèle probabiliste <u>dans le cadre de la théorie des probabilités Θ_p</u>, où Θ_p inclut les théories Θ_{mi} précédemment définies : ce cadre est celui qui est proposé pour le cycle seconde / première S

---

[85] hypothèse qu'autorisent les diagrammes en batons représentant la série des valeurs observées.

/ terminale S dans les projets de programme 2010.

Nous présentons dans le tableau ci-dessous les éléments théoriques correspondant aux praxéologies ponctuelles décrites précédemment et Θ_p qui peut être prise comme *théorie globale* au sens de Chevallard.

L'illustration précédente de la résolution du problème de M. Dupond ne nécessite en fait que la technologie θ_{m2} et la définition de l'intervalle de confiance $[E(Y) - 2\sigma ; E(Y) + 2\sigma]$ au niveau 95 % (résultat qui, curieusement, ne figurait qu'au programme 2000 du tronc commun de première L, et qui figure dans les programmes 2010 de la classe de seconde)

Θ_{m1}	Θ_{m2}	Θ_{m3}	Θ_p
- espace probabilisé fini $(\Omega ; \mathcal{P}(\Omega) ; p)$; espérance et écart type d'une loi ; variable aléatoire, espérance mathématique et écart type de la variable. - analyse combinatoire : p-listes, permutations, combinaisons	Θ_{m1} et les probabilités conditionnelles, et la formule des probabilités totales, qui justifie les règles de calcul dans un arbre de probabilité. NB : la notion d'indépendance stochastique n'est pas nécessaire.	espace probabilisé fini $(\Omega ; \mathcal{P}(\Omega) ; p)$ et loi des grands nombres (loi empirique, ou théorème de la théorie des probabilités) ; la convergence des fréquences est une convergence en probabilité.	- espaces probabilisés, finis ou infinis (lois à densités continues). - loi des grands nombres. - variable aléatoire, espérance mathématique et écart type. - probabilités conditionnelles, événements indépendants. - lois binomiales, lois normales, etc. (« programmes 2010 » de terminale non fixé). - intervalles de confiance et niveau de confiance.

4. L'utilisation du modèle informatique

Nous avons illustrée cette utilisation du modèle informatique dans le chapitre précédent, avec la résolution du problème de M. Dupond.

Cette utilisation peut se faire dans le seul cadre de l'algorithmique et du calcul d'indicateurs de statistique descriptive (calcul de centiles, de moyenne et d'écart-type), dans le cas (fréquent en lycée, peut-être même est-ce toujours le cas) où les *hypothèses de modélisation* concernant les expériences élémentaires en jeu relèvent des expériences aléatoires paradigmatiques déjà évoquées .

La simulation doit permettre de calculer une (ou plusieurs) variable aléatoire et doit être répétée un grand nombre de fois, pour permettre de récupérer une série statistique par variable aléatoire, et de faire des calculs de statistique descriptive sur ces séries.

La mise en œuvre informatique de ces répétitions et de ces calculs statistiques relève aussi de tâches de type algorithmique, élémentaire pour ce qui concerne la répétition un nombre n donné de fois de la simulation, moins élémentaire pour les calculs statistiques, mais qui peuvent relever de *boîtes noires* (procédures prises dans une bibliothèque de sous-programmes, ou mise en œuvre des fonctions statistiques du tableur ou de la calculatrice).

Nous avons vu que cette utilisation peut également être placée dans le cadre théorique $\Theta_p \cup \Theta_s$.

Nous entrons alors dans une *organisation globale* pour l'institution I_T qu'est la classe de terminale avec les programmes 2010.

L'objet d'étude (*le système réel*) est une situation où intervient une expérience aléatoire E.A.

Les tâches possibles sont la modélisation probabiliste et la modélisation informatique de cette expérience E.A.

Nous allons voir dans le chapitre 22 quels objectifs didactiques on pourrait assigner à ces tâches, tout particulièrement à la simulation informatique, pour développer ce que d'aucuns appellent *l'esprit statistique* ou *l'esprit probabiliste*.

Retenons que la simulation de type *imitation au plus près* d'une expérience aléatoire est peut-être l'une des techniques évoquées par Chevallard (1999) lorsqu'il écrit : « *il existe des technologies* ***potentielles****, en attente de techniques, qui ne sont encore technologies d'aucune technique ou de très peu de techniques* ».

CHAPITRE 21

A propos de l'esprit probabiliste

1. « Esprits »

Blaise Pascal (1670) dans « Les pensées », parlait de *l'esprit de géométrie* et de *l'esprit de finesse* (et aussi de *l'esprit de justesse*) en tentant de préciser des démarches intellectuelles qui lui semblaient spécifiques à des domaines particuliers.

On retrouve des formulations analogues pour caractériser des démarches intellectuelles qui seraient spécifiques à des domaines particuliers : *l'esprit scientifique* chez Bachelard (1938), *l'esprit algorithmique* (Kahane (2000)), *l'esprit statistique* et *l'esprit probabiliste* (nombreux auteurs comme nous allons le voir), « esprits » qu'il faudrait introduire ou renforcer dans l'enseignement.

1.1. L'esprit scientifique

Dans son ouvrage *La formation de l'esprit scientifique*, Bachelard ne définit pas en fait un *esprit scientifique*, mais essaye plutôt, en épistémologue des sciences, et, dans cet ouvrage, plus particulièrement des sciences expérimentales [86], de saisir différentes périodes et étapes qui ont mené à ce qu'il appelle *l'ère du nouvel esprit scientifique*, après *l'état préscientifique* et *l'état scientifique* (il situe ces périodes de l'antiquité classique au début du XVIII^{e} siècle, y incluant la Renaissance, puis du XVIII^{e} au début du XX^{e} siècle, datant la dernière période de la Relativité d'Einstein) :

> *« Mais nous ne nous astreindrons pas à inscrire nos remarques particulières dans ce triptyque qui ne nous permettrait pas de dessiner avec assez de précision les détails de l'évolution psychologique que nous voulons caractériser. Encore une fois, les forces psychiques en action dans la connaissance scientifique sont plus confuses, plus essoufflées, plus hésitantes, qu'on ne se l'imagine quand on les mesure du dehors, dans les livres où elles attendent le lecteur. (...) Même chez un esprit clair, il y a des zones obscures, des cavernes où continuent à vivre des ombres. Même chez l'homme nouveau, il reste des vestiges du vieil homme. (...) »*

---

[86] Bachelard considère qu'il convient de mettre les mathématiques à part, nous reviendrons plus loin sur ce point de vue.

Bachelard décrit à la fois les conditions du développement de la pensée scientifique et de la pensée du scientifique :

> *« Quand on cherche les conditions psychologiques des progrès de la science, on arrive bientôt à cette conviction que* ***c'est en termes d'obstacles qu'il faut poser le problème de la connaissance scientifique****. Et il ne s'agit pas de considérer des obstacles externes, comme la complexité et la fugacité des phénomènes, ni d'incriminer la faiblesse des sens et de l'esprit humain : c'est dans l'acte même de connaître, intimement, qu'apparaissent, par une sorte de nécessité fonctionnelle, des lenteurs et des troubles. C'est là que nous montrerons des causes de stagnation et même de régression, c'est là que nous décèlerons des causes d'inertie que nous appellerons des obstacles épistémologiques. (...) En fait, on connaît* ***contre*** *une connaissance antérieur, en détruisant des connaissances mal faites, en surmontant ce qui, dans l'esprit même, fait obstacles à la spiritualisation. »*

Cette notion d'*obstacle épistémologique* peut être, dit Bachelard, étudiée dans le développement historique de la pensée scientifique, et aussi <u>dans la pratique de l'éducation</u>. Or dit-il, cette notion d'obstacle épistémologique est méconnue dans l'éducation :

> *« Les professeurs de sciences imaginent que l'esprit commence comme une leçon, qu'on peut toujours refaire une culture nonchalante en redoublant une classe, qu'on peut faire comprendre une démonstration en la répétant point par point. Ils n'ont pas réfléchi au fait que l'adolescent arrive dans la classe de Physique avec des connaissances empiriques déjà constituées : il s'agit alors, non pas* ***d'acquérir*** *une culture expérimentale, mais bien de* ***changer*** *de culture expérimentale, de renverser les obstacles déjà amoncelés par la vie quotidienne. »*

Nous reviendrons sur ce dernier point qui semble illustrer l'une des difficultés de l'enseignement de la statistique et des probabilités.

<u>1.2. L'esprit algorithmique</u>

Dans un rapport J.P. Kahane (2000) définit *l'esprit algorithmique* :

> *« La plupart des machines, ordinateurs ou robots, fonctionnent avec des programmes qui agencent la succession désirée de tâches élémentaires. Ces programmes mettent en jeu des algorithmes traduits en langage compréhensible*

par la machine. (…)
Ces algorithmes, même les plus simples, mettent en jeu des structures de données comme les arbres ou les graphes dont les sommets représentent les différents états du système en fonctionnement. De même que chaque individu est plus ou moins familier avec l'espace qui l'entoure et capable de s'y repérer, de s'y mouvoir, d'en concevoir la globalité et le détail, ce même individu se repère, se déplace, utilise ces structures et rend compte de leur utilisation avec plus ou moins de familiarité.

L'analogie avec la perception de l'espace n'est pas fortuite puisque les graphes, les arbres sont symbolisés par des diagrammes.

Décomposer une tâche complexe en tâches élémentaires, reconnaître les tâches qui se répètent ou qui ont été déjà traitées, estimer la durée du processus, sans oublier de vérifier que la succession d'opérations élémentaires produit bien le résultat escompté : c'est la démarche de celui qui écrit un programme, c'est aussi celle de tout être rationnel.

Dans l'enseignement ***l'esprit algorithmique*** [87] *se manifeste assez tôt comme objet d'enseignement, par exemple lors de la résolution des premiers problèmes ''d'arithmétique élémentaire'' comportant plusieurs ''calculs'' ; par la suite, il accompagne les résolutions et les démonstrations mathématiques à tous les niveaux. Les solutions sont traditionnellement présentées sous forme de ''raisonnements'' à l'école primaire, mais qui n'ont pas de statut solide par la suite, de sorte que leur formulation tend à disparaître, même aux niveaux supérieurs. Ces solutions pourraient être plus commodément cherchées, étudiées, établies, explicitées, apprises et enseignées grâce à un usage concerté (au moins reconnu tout au long de la scolarité) des instruments de l'algorithmique. »*

J.P. Kahane essaye dans ce rapport d'étape sur l'informatique et l'enseignement des mathématiques de convaincre que l'enseignement des mathématiques doit prendre en charge l'algorithmique-programmation, et développe divers arguments, en particulier celui d'un *esprit algorithmique* qui accompagnerait la démarche de tout être rationnel.

1.3. L'esprit probabiliste et l'esprit statistique

Georges Darmois (1888-1960) était un précurseur dans l'enseignement des statistiques ; il fut

[87] Souligné par nous.

directeur de l'Institut de Statistiques de l'Université de Paris. Dans la *Revue de Statistique Appliquée*, à propos de « la formation probabiliste et statistique de l'ingénieur », il décrit deux réactions possibles devant une pièce produite par un tour, et qui ne serait pas au diamètre attendu : l'une liée à *une conception déterministe*, l'autre liée à *l'esprit statistique*.

Il explique alors :

> *« Au fond, c'est l'élargissement de l'esprit de recherche, et l'Ingénieur ne doit pas être seul à le posséder. Il faut que, par son action et son influence, pénètrent de nouvelles formes de penser et de réagir. c'est la richesse et la qualité des résonances nouvelles ainsi crées qui permettra de résoudre des questions qu'on n'avait même pas aperçues auparavant. Cet enrichissement doit être recherché pour les jeunes et pour les adultes. »*

Il poursuit :

> *« Et, insistant maintenant sur ce qui est mon but, je voudrais dire qu'il y a beaucoup plus que l'avantage immédiat de certaines applications, c'est justement l'élargissement mental auquel conduit* ***l'esprit probabiliste et statistique****.*[88]
> *Tout le monde connaît bien cette impression d'élargissement que donne une théorie nouvelle, expliquant tous les résultats des théories précédentes, c'est à dire les comprenant comme cas particuliers ou cas limites, mais enrichie de points de vue plus élevés, expliquant par là ce qui échappait aux théories anciennes, et construisant par la vie normale de la pensée des idées nouvelles, des possibilités entièrement neuves de réalisations.*
> *Je voudrais faire une brève revue de ces nouveautés dont quelques-unes résultent d'une longue mise au point, comme la théorie des probabilités elle-même, qui débutait il y a environ 3 siècles, mais qui, reconnaissons-le, n'est pas encore bien entrée dans le système de l'instruction publique (Laplace, un siècle et demi). »*
> Darmois (1956)

Et Darmois donne alors des exemples de théories et d'outils nouveaux apportés par la statistique et les probabilités : *la théorie des échantillons*, et la notion d'*écarts significatifs* (tests d'hypothèses), la notion de *dépendance stochastique* (corrélation), l'idée de *plan expérimental* :

> *« Une grande idée de Jacques Bernoulli est qu'on pouvait conduire vers la connaissance un cheminement aléatoire. La façon de conduire ce cheminement est ce qu'on peut appeler de façon générale le plan d'expérience (expérimental*

[88] souligné par nous.

design). C'est là une dignité nouvelle du statisticien moderne. Associé à la recherche dès son début et dans ses pas successifs, il est chargé d'indiquer comment obtenir l'information la plus étendue avec certains moyens.
Ce cheminement optimum vers la connaissance peut être aidé et repéré par la théorie des informations. Il peut s'arrêter quand on juge l'information suffisante pour la décision (cheminement séquentiel, estimation la meilleure, estimation exhaustive de l'information obtenue…). »

Darmois décrit enfin les application variées du *concept d'évolution aléatoire* (processus stochastiques) : évolution d'une population (au sens large) en économie, physique, démographie.

Nous avons trouvé des références à *l'esprit statistique* ou à *l'esprit probabiliste*, entre autres, chez Fischbein (cf. le paragraphe 4 de ce chapitre), dans les documents d'accompagnement des « programmes 2000 » de la classe de seconde (p. 7) :

« L'esprit statistique naît lorsqu'on prend conscience de l'existence de fluctuation d'échantillonnage »,

et chez Régnier (2003) :

« La formation de l'esprit statistique » s'inscrit dans une dynamique qui permet aux être humains de se confronter aux obstacles – épistémologiques ou autres – au développement de la statistique en tant que science. »

Coutanson (2010) consacre une partie de sa thèse à définir *le fait statistique, la pensée statistique,* et *l'esprit statistique*.

2. Un relatif échec de l'enseignement

On constate un certain échec de l'enseignement dans ces domaines de la statistique et des probabilités, échec que décrit Lahanier–Reuter [1999] et particulièrement par cette phrase que nous avons déjà citée : « *L'intérêt de la modélisation de la situation est à reconstruire par l'élève, puisque la synthèse des informations nouvelles que permet cette modélisation n'est jamais proposée, ni demandée* ».

Courtebras (2008) confirme ce constat en analysant les programmes de BTS, et les sujets d'examens correspondant :

« L'enseignement probabiliste destiné aux élèves techniciens supérieurs consiste essentiellement à porter à leur connaissance le vocabulaire spécifique au calcul des probabilités et aux phénomènes aléatoires, à leur apprendre à faire fonctionner,

dans les exercices stéréotypés d'application, les lois de probabilité les plus usuelles – loi binomiale, loi de POISSON, loi normale – et à savoir calculer une probabilité conditionnelle. (…) »

Il semble que l'un et l'autre pointent une même difficulté : les élèves / étudiants ne sont pas placés dans une situation appelant une prise de décision pour une action dans des conditions incertaines, prise de décision pour l'action que devrait éclairer une étude scientifique de ces conditions incertaines, de type statistique ou probabiliste.
Lahanier met en cause le non retour à la situation réelle de départ donnant un sens au modèle probabiliste calculé ; Courtebras met en cause les phénomènes aléatoires mêmes que l'on donne à étudier et la forme des questionnements proposés.

En lycée, le *modèle probabiliste* produit des connaissances nouvelles sur l'expérience aléatoire, mais, en dehors de situations relativement élémentaires [89], il faut des connaissances en théorie des probabilités qui vont bien au delà de la notion d'espace probabilisé pour pouvoir exploiter ces connaissances, ces informations nouvelles comme dit Lahanier-Reuter.

Ces connaissances supplémentaires mettent en jeu des notions mathématiques a priori extérieures au cours de probabilité : outils de dénombrement pour les lois de Bernouilli ou les lois de Poisson, calcul intégral pour les lois à densité continue (uniforme, exponentielle, normale), notions difficilement envisageables avant la classe de terminale, et encore serait-il peut-être risqué d'enchaîner la même année ces notions de probabilité à la suite du cours d'intégration.
Il en va ainsi du théorème de Moivre-Laplace et de ses applications aux *intervalles de fluctuation*, *intervalles de confiance* et *niveau de confiance*, qui figurent dans les projets de programme pour la terminale S (« programmes 2010 »).

On peut imaginer de placer dans les programmes des formules qu'on ne démontrerait pas permettant par exemple de calculer ces intervalles. Encore faudrait-il que, au delà de la capacité ou non des élèves à déterminer par le calcul ces intervalles, les notions même qui leur sont sous-jacentes soient comprises par les élèves.

Le bilan que l'on pourrait faire de l'enseignement des probabilités, tel qu'il se pratique semble-t-il encore avec les programmes 2000, reste dans la description de Lahanier-Reuter déjà évoquée : une sorte de propédeutique pour former peut-être à *l'esprit probabiliste*, et pourrions-

[89] Les situations de prise de décisions telles que décrites par Maury (1986), ou celles décrites par Boulanger (2009) qui peuvent dans le cadre de la classe être résolue avec un calcul d'espérance et d'écart-type : ces situations peuvent servir à légitimer ces notions.

nous même dire, il reste dans *l'esprit géométrique* : il y a des hypothèses et des axiomes, des théorèmes, et on fait fonctionner la règle du *modus ponens*.

Les concepteurs des « programmes 2000 », en voulant insister dans l'ensemble de la prescription [90] sur *l'activité modélisatrice* du scientifique, ont peut-être sur-privilégié dans le programme de probabilité l'aspect relatif du modèle, lié aux hypothèses de modélisation, et ce au détriment de l'exploitation du modèle, et partant, peut-être, de *l'esprit probabiliste*.

Le *modèle informatique* produit aussi des informations nouvelles sur l'expérience aléatoire, que l'on peut exploiter sans connaissances en théorie des probabilités en dehors des notions d'espace probabilisé et de variable aléatoire, et de la loi des grands nombre*s, constatée* sur des expériences aléatoires, simulées ou non.
Nous pensons l'avoir montré avec l'exemple du problème de M. Dupond.

Nous allons tenter de cerner ce qui, peut-être, dans l'exploitation de cet autre type de modèle d'une expérience aléatoire, pourrait être un outil dans la formation à ce que certains donc appellent *l'esprit probabiliste*.

3. Y a-t-il un *esprit probabiliste* ?

Cette expression « esprit probabiliste » nous a étonnée lorsque nous l'avons rencontrée pour la première fois, même si nous avions en tête *l'esprit géométrique* et *l'esprit de finesse* de Pascal, et l'*esprit scientifique* de Bachelard.
Cependant, force nous est de constater qu'il manque quelque chose dans l'enseignement des probabilités et peut être dans l'enseignement de la statistique (mais est-ce la même absence ?), et qui ne se résume pas à une absence de formation des enseignants : les enseignants ont su intégrer assez vite la théorie des ensembles [91], l'algèbre linéaire et d'autres domaines qu'ils n'avaient peut-être pas rencontrés dans leur parcours universitaire [92].

Les réticences qui s'expriment encore dans les années 2010, après une décennie de l'enseignement des « programmes 2000 », sur la nécessité d'un enseignement des probabilités et de la statistique au lycée, au travers par exemple des forums évoqués dans la première

---

[90] Que ce soit en géométrie, analyse, probabilités (mais c'est peut-être en probabilités que c'était le plus facile) pour les mathématiques, mais également en sciences physiques et en SVT, dans un souci d'affirmer l'unité de la démarche scientifique.

[91] cela n'occulte pas les ravages dans l'enseignement qui ont pu en résulter.

[92] tel que la géométrie dans l'espace pour les « enfants » des maths modernes.

partie, rapprochées de l'optimisme affiché il y a plus de cinquante ans par Darmois et Fréchet concernant l'introduction d'un enseignement de la statistique et des probabilités dans l'enseignement secondaire (cf. la fin de l'article cité au paragraphe 1 de ce chapitre), alors même que ces domaines des mathématiques continuent de se développer, peuvent laisser penser qu'il ne s'agit pas que d'une absence de formation dans ces domaines [93], mais peut être aussi d'une mauvaise formation.

Car après tout, est-ce que les enseignants eux-mêmes, dans leur formation, ont dépassé ce qui est actuellement demandé aux élèves de lycée, résumé par le constat, déjà évoqué, de Lahanier-Reuter : une fois modélisée la situation aléatoire, il n'est pas proposé à l'élève de faire quelque chose des informations nouvelles sur la situation qu'apporte le modèle qui vient d'être élaboré ?
Pour dire les choses autrement, la formation des enseignants [94] aux probabilités n'est-elle pas essentiellement une formation à la théorie des probabilités au sens de Kolmogorov (dans une certaine mesure une formation très *Bourbakiciste* [95]) dans le cadre de la théorie des ensembles et éventuellement de la théorie de la mesure, d'une part occultant les questionnements historiques et épistémologiques sur le hasard, les probabilités, et leur applicabilité, et d'autre part laissant de côté la confrontation de la théorie avec *de vrais problèmes* à résoudre dans des situations incertaines ?
Et on peut poser la même question pour la formation des enseignants à la statistique.

Nous n'avons pas ici l'ambition de faire une étude raisonnée des idées exprimées par les enseignants de lycée sur les contenus souhaitables des programmes dans le domaine des statistiques et des probabilités. Nous avons extrait deux phrases d'un échange en date du 14 avril 2011 sur le forum du site de l'APMEP, à propos des projets de programme pour la terminale S (rentrée 2012) soumis à la consultation des enseignants, concernant les probabilités et la statistique :

« (...) Un saut qualificatif bien trop important dans ce chapitre - notamment dans

---

[93] absence réelle pour certains enseignants.

[94] que ce soit leur formation initiale (lycée et post-bac) et leur formation d'enseignant, ici leur formation pour enseigner les probabilités ou la statistique.

[95] à rapprocher des dérapages qui ont pu amener, dans les programmes dits des maths modernes, à présenter en lycée la géométrie par l'algèbre linéaire, une rotation du plan étant définie comme une transformation affine associée à une transformation orthogonale positive.

les lois à densité. Les lycéens n'ont presque vu que des cas d'équiprobabilité.
Il faut simplifier de façon très significative cette partie de programme sous peine de mettre en échec un grand nombre d'élèves. Un très faible nombre aura véritablement besoin de notions aussi précises sur ce chapitre, et ils auront tous le temps de l'étudier. Cela va prendre beaucoup de temps pour un apport réduit dans leur formation intellectuelle. Il y a des chapitres dans lesquels on peut approfondir. On peut économiser du temps ici pour l'investir ailleurs.(...)». B.

« (...) Un peu de courage que diable osons dire que leur informatique et leurs statistiques n'ont rien de formateur !
Dites aux inspecteurs qu 'ils créent leur épreuve d'informatique garantie pure beurre, qu'ils en fassent leur élément fondamental de sélection.
Dites aux inspecteurs que nous, travailleurs de basse souche, comprenons parfaitement leur aliénation aux avantages de la position, et qu'en conséquence nous ne leurs en voudrons pas de créer une option statistique complètement inutile en terme de formation et de construction individuelle, mais qui leur permettra de préserver ces petits plaisirs qui font que la vie mérite d'être vécue selon certains (...) ». L.G.

Nous avons fait ressortir ces deux extraits car ils nous semblent exprimer précisément la négation de toute spécificité à la statistique et aux probabilités quant aux démarches à mettre en œuvre, et à la difficulté d'acquisition de ces démarches.
D'autres contributions trouvées sur ce site sont moins négatives, tout en soulignant la difficulté à enseigner les notions proposées, particulièrement pour des enseignants insuffisamment formés, mais aucune participation sur ce site n'évoque une nécessaire formation à un *esprit statistique* ou *probabiliste*, des élèves et/ou des enseignants, alors que fréquemment sont soulignés les apports de la géométrie à la formation individuelle / scientifique de l'élève, avec le regret que la géométrie des nouveaux programmes soit surtout de *la géométrie repérée*.
Les projets de programmes de terminales S pour la rentrée scolaire 2012 ne sont pas particulièrement explicites quant à la prise en compte d'une démarche spécifique en statistiques et probabilités : si des notions nouvelles apparaissent, telles que la notion d'intervalle de fluctuation, elles sont présentées dans des termes qui peuvent paraître surtout techniques.

On peut consulter en annexe de ce chapitre les tableaux qui définiraient le programme de statistique et de probabilités en terminale S si ces programmes étaient adoptés.
Notons que « statistiques » dans le titre est encore [96] au pluriel.
Remarquons également que si le titre de cette partie du programme est « Probabilités et statistiques », il reste difficile dans ce qui est proposé de distinguer les rubriques qui relèvent de la statistique de celles qui relèvent des probabilités. En fait, il semble bien que tout se ramène au domaine des probabilités, à l'exception de l'alinéa « *La simulation de sondages sur tableur permet de sensibiliser aux fourchettes de sondage* », qui en l'absence de documents d'accompagnement qui préciseraient cette déclaration plutôt sibylline, laisse entière la possible confusion entre les deux domaines au travers de l'ambiguïté dans ce contexte du terme *échantillon*.
Soulignons enfin le chapeau d'introduction :

> « *On approfondit le travail en probabilités et statistiques mené les années précédentes.*
> *Afin de traiter les champs de problèmes associés aux données continues, on introduit les lois de probabilités à densité. Le programme en propose quelques exemples et, en particulier, la loi normale qui permet notamment d'initier les élèves à la statistique inférentielle par la détermination d'un intervalle de confiance d'une proportion à un niveau de confiance de 95 %.*
> *Cette partie se prête particulièrement à l'étude de problèmes issus d'autres disciplines.*
> *Le recours aux représentations graphiques et aux simulations est indispensable.* »

La nouveauté de ce projet semble résider dans l'introduction des notions d'*intervalle de confiance*, associé à un *niveau de confiance*, mais dans une formulation qui semble plutôt théorique, qui légitime certes l'introduction de la notion de loi de probabilités à densité, nécessaire pour introduire la loi normale, elle-même en apparence nécessaire pour introduire ces intervalles et niveaux de confiance.
Mais ce qui est visé réellement n'apparaît pas clairement : s'agit-il de notions nouvelles permettant des prises de décision en situation d'incertitude (*intervalle de confiance* et *niveau de confiance* associé) ou d'applications possibles de la nouvelle notion de *loi de probabilité à*

[96] Comme dans les programmes antérieurs.

densité continue [97] ?

Le fait que les items de cette partie du programme semblent plutôt articulés autour de démonstrations de résultats d'analyse, appliqués ensuite à une nouvelle problématique, spécifique des probabilités, les *intervalles* et *niveaux de confiance*, fait craindre que la mention laconique du chapeau d'introduction :

> *« Cette partie se prête particulièrement à l'étude de problèmes issus d'autres disciplines.*
>
> *Le recours aux représentations graphiques et aux simulations est indispensable. »,*

ne reste lettre morte.

Le « programme 2010 » de la classe de seconde, actuellement en vigueur, que l'on peut trouver en annexe du chapitre 3, introduit la notion d'intervalle de fluctuation d'une fréquence au seuil 95 %. On ne sait pas si le paragraphe, qui a pour titre « échantillonnage », relève de la statistique ou des probabilités. Le mot « simulation » à trois reprises apparaît sans complément, et une fois dans la locution : « simulation de situations concrètes ».

On lit dans la colonne « commentaires »

> *« L'objectif est d'amener les élèves à un questionnement lors des activités suivantes :*
>
> *• l'estimation d'une proportion inconnue à partir d'un échantillon*
>
> *• la prise de décision à partir d'un échantillon » .*

Une note est donnée en bas de page, destinée probablement à préciser à l'enseignant :

- la définition d'un intervalle de fluctuation au seuil 95 %,
- que « *cet intervalle peut être obtenu, de façon approchée, par simulation* » ,
- et la formule $[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}]$ que « *le professeur peut indiquer aux élèves* », avec la précision suivante :

> *« le professeur peut faire percevoir expérimentalement la validité de cette*

[97] Les lois de probabilité à densité continue, version programmes 2000, étaient limitées aux lois uniformes sur [a ; b] et à la loi exponentielle qui, une fois démontrée qu'une loi à durée de vie sans vieillissement est une loi exponentielle (démonstration difficile pour les élèves) ; l'application de ce résultat à la désintégration nucléaire en sciences physiques et en SVT faisait partie des « programmes 2000 » de ces disciplines ; en revanche ce résultat ne débouchait pas en classe de mathématiques sur des exercices très significatifs en dehors de contextes de circuits électroniques, dont la durée de vie serait sans vieillissement.

propriété mais ***elle n'est pas exigible*** *»* (B.O., idem)

Il semble qu'il y a certainement des façons d'enseigner ce qui vient d'être évoqué (concernant la classe de seconde, et le troisième paragraphe du projet de programme de terminale S), en occultant qu'il s'agit de *mathématiques appliquées*, ou plutôt de *mathématiques applicables* à des problèmes spécifiques, nécessitant peut-être une démarche particulière que certains appellent l'*esprit probabiliste*.
Et sans cet *esprit*, nous sommes nous aussi tentée de dire que cette prescription risque pour les élèves d'être un « *apport réduit dans leur formation intellectuelle »* comme nous l'avons trouvé écrit sur le site de l'APMEP déjà cité.

4. Qu'est-ce que *l'esprit probabiliste* ?

Mais finalement, qu'est-ce que *l'esprit probabiliste* ou *l'esprit statistique* ? Nous n'avons pas l'intention de distinguer deux démarches (même si les champs de la statistique et des probabilités ne sont pas les mêmes), ni de trancher sur la pertinence de ces deux locutions.
Ce qui nous importe ici, c'est de voir si les simulations informatiques d'expériences aléatoires, modèles alternatifs en même temps que complémentaires aux modèles classiques probabilistes, peuvent être utilisées pour éclairer des notions qui paraîtraient centrales dans ces domaines que sont les probabilités et la statistique.

Nous allons donc reprendre les idées forces que nous avons trouvées chez certains auteurs, autour de cette hypothèse qu'il y aurait des démarches spécifique aux probabilités et à la statistique, et essayerons d'en faire une synthèse, modeste, en lien avec notre recherche sur les simulations informatiques.

4.1. Fischbein

Fischbein, Pampu et Minzat (1969) écrivent, à propos de l'introduction en lycée d'un enseignement de la statistique et des probabilités :

> *« Mais l'expérience a montré que les adultes se heurtent à des grandes difficultés même s'il s'agit de chapitres élémentaires. En réalité, les probabilités ne représentent pas seulement un ensemble de techniques de calcul, mais* ***un mode de pensée spécifique****, exprimant une orientation distincte de l'intellect (Servais, 1968). Ceci signifie que l'introduction de certaines notions de probabilité et de statistique doit commencer tôt et, évidemment, sous forme élémentaire, de*

manière que l'orientation respective soit cultivée graduellement, en mettant en valeur les potentialités intellectuelles des enfants et en les exerçant par des moyens appropriés, ainsi qu'on procède avec l'arithmétique, l'algèbre ou la géométrie.(...) »

Ils précisent que si l'idée de probabilité, en terme de chance apparaît après 7 ans (il s'agit, en reprenant Piaget et Inhelder (1951) de comprendre l'opposition entre le déterminé et l'indéterminé au niveau des opérations concrètes), et si, toujours selon Piaget et Inhelder, la capacité à utiliser le concept de probabilité ne peut se faire qu'au stade des opérations formelles, il faut cependant prendre en compte le fait que

> *« l'évolution mentale ne prépare, en réalité, qu'une série de potentialités. Leur mise en valeur effective ne peut se réaliser que par un exercice systématique, de longue durée, au cours de ces stades de l'évolution intellectuelle. »*

En d'autres termes, ces potentialités pourraient bien ne jamais se manifester si les enfants ne sont pas préparés, et avant même qu'elles puissent s'exercer, à rencontrer des situations où elles pourront s'exercer.

Fischbein, Pampu et Minzat [1967] ont en effet monté une expérience tendant à montrer que pour certaines questions, les enfants semblaient avoir régressé au cours de leur scolarité :

> *« Mais pourquoi cette dégradation surprenante des réponses chez les grands élèves ? L'explication est selon nous la suivante. Le processus d'enseignement – surtout par l'intermédiaire de l'école – oriente l'enfant vers une interprétation déterministe des phénomènes, dans le sens de la recherche et de l'explicitation de relations claires, certaines, univoques. L'enfant est instruit à chercher les causes des phénomènes sous la forme de facteurs agissant d'une manière univoque. Théoriquement, les concepts de nécessité et de hasard, de certitude et d'incertitude se délimitent et s'équilibrent réciproquement dans la dynamique de l'activité mentale. Mais en réalité, dans l'évolution intellectuelle de l'enfant, la délimitation et l'équilibration sont profondément perturbées du fait que* ***l'instruction scolaire ne favorise systématiquement qu'un seul de ces termes****. »*

Et plus loin :

> *« Au fond, l'enfant arrive graduellement à être convaincu du fait que comprendre pleinement un phénomène, signifie le rapporter à une relation causale univoque, croyance qui a dominé, en réalité, la pensée scientifique moderne pendant des*

siècles et que l'école cultive toujours. »

Les auteurs pensent le problème « *en principe* » résolu par l'introduction du calcul des probabilités dans les grandes classes, et préconisent l'introduction de problèmes pratiques et théoriques même pour les petits élèves.

4.2. Darmois

Dans un article précédemment cité, Darmois (1956) décrit deux attitudes possibles devant une situation concrète :

> *« Nous attendons un résultat, par exemple, que sorte une pièce d'un certain diamètre. Il sort une pièce de dimension plus forte. Que faire ? Si notre conception est, disons déterministe, nous croyons qu'à un réglage donné correspond un diamètre déterminé ; la conclusion s'impose : il faut changer le réglage. Si, au contraire, nous concevons le résultat comme aléatoire, si nous savons qu'un tour fabrique des pièces fluctuantes, nous devons, non pas attendre passivement, mais observer, mesurer les fluctuations, leur distribution, la stabilité de cette distribution.*
>
> *Ce que certains croient être un changement vrai nécessitant un changement de politique et d'organisation, n'est souvent qu'un changement aléatoire qu'il faut étudier numériquement pour conclure si l'organisation actuelle est vraiment adaptée à nos desseins.*
>
> *Le tour, par exemple, peut être censuré ou être utilisé à d'autres besognes moins fines.*
>
> *C'est dans des cas de ce genre qu'apparaît ce que je crois être la supériorité de l'esprit statistique. »*

Darmois dans son plaidoyer pour un enseignement de l'*esprit statistique*, envisageait donc une opposition possible, dans des pratiques professionnelles, entre une démarche déterministe, et une démarche prenant en compte l'aléatoire, opposition que cette formation devrait résoudre.

4.3. L'esprit statistique vu par Régnier

Régnier ((2002) et (2005)) essaye de définir l'*esprit statistique*. Pour cela il définit les objectifs de la formation :

« (…) la formation en statistique vise à rendre l'apprenant capable de :

• construire un modèle requis par une approche statistique congruent au modèle dans lequel une problématique est posée.

• décrire, traiter, analyser des données de manière pertinente dans le cadre du modèle construit pour conduire une étude.

• tenir un raisonnement intégrant l'idée de risque dans l'énoncé des conclusions.

• interpréter les phénomènes étudiées sur la base de données statistiques recueillies sur des faits et leurs relations.

• communiquer des résultats des analyses de données en faisant une distinction entre le modèle utilisé et la réalité qu'il est supposé représenter, entre les traitements conduits au sein du modèle et les interprétations reformulées dans le contexte au sein duquel est posée la problématique.

• poursuivre de façon autonome et personnalisée un apprentissage en statistique afin d'enrichir les acquis personnels actuels ». Régnier [2002].

Et plus loin :

*« Nous postulons que ni l'**esprit statistique** n'est un don, ni sa formation n'est le résultat du développement naturel de l'être humain. (Régnier 1998 b). La formation de l'**esprit statistique** se caractérise par la nature du rapport à l'incertitude et à l'erreur, considérées comme inhérentes à tout acte de prise de décision. Elle est instrumentée par une conceptualisation du risque encouru dans une prise de décision et par une modélisation de son contrôle, ne laissant plus l'exclusivité à une compréhension ou une explication fondées sur une conception spontanée du hasard et du déterminisme, empreinte parfois du fatalisme. »*

4.4. Maury et les probabilités

Dans sa thèse, Maury (1986), se référant aux travaux de Fischbein (déjà cités) et de Brousseau (1971) et s'étonnant qu'on n'enseigne pas les probabilités plus précocement, souligne leur spécificité :

« en effet les probabilités ne se réduisent pas seulement à un ensemble de "techniques de calcul" et certains considèrent qu'elles constituent un mode de pensée spécifique nécessitant la création précoce d'un corps "d'habitudes mentales". En définitive, l'intérêt accordé par l'école à l'orientation des programmes d'enseignement dans le sens du développement scientifique, en tenant compte de l'importance croissante des probabilités dans la pensée scientifique est relativement faible.(…) »

La statistique et les probabilités sont des parties du programme souvent délaissées :

> *« L'importance manifeste que présentent ces disciplines* [98]*- pour la formation individuelle des élèves, mais aussi pour faciliter leur passage à la vie active ou à l'université - incite à s'interroger sur cette situation.On peut certes invoquer, pour l'expliquer, l'insuffisance (voire l'absence) de formation des maîtres dans ce domaine (y compris des enseignants de l'université). On peut aussi invoquer la pauvreté de l'expérience pédagogique collective (…)*
> *L'enseignement d'une nouvelle branche de la science soulève toujours de nombreuse questions – en particulier d'ordre cognitif et didactique – que l'on ne peut pas résoudre à l'intérieur de la science elle-même ou de son histoire ni réduire à un manque de formation des enseignants. En ce sens, l'une des données décisives, en ce qui concerne la situation que nous avons exposée, touche à la rareté des réflexions didactiques sur le sujet, réflexions relevant d'études systématiques des programmes et des procédures d'enseignement et fondées sur la confrontation entre les processus ou les raisonnements exigés par ces disciplines et les activités cognitives réelles des élèves. »*

Maury remarque que la plupart des travaux concernent les jeunes enfants ou les pré-adolescents.

> *« (…) en revanche, l'on dispose de peu de résultats concernant l'élaboration des connaissances propres à ce domaine chez l'adolescent et le jeune adulte scolarisés alors que c'est précisément cette population qui est concernée par l'enseignement de la statistique et des probabilités.*
> *A cela il faut ajouter que, de l'avis des enseignants ayant en charge ces disciplines, élèves et étudiants éprouvent des difficultés particulières à leur abord. (…) de ce point de vue la situation de la statistique et des probabilités est analogue à celle de la mécanique classique, domaine qui a fait l'objet d'importants travaux didactiques (L. Viennot 1979, E. Saltiel 1978). Cette analogie pourrait provenir de ce que* ***la statistique et les probabilités, tout comme la mécanique (et la physique plus généralement) entretiennent des liens étroits avec la réalité. Elèves et étudiants aborderaient alors l'apprentissage scolaire de ces disciplines avec un corps de connaissances spontanées important (cf : les modèles spontanés chez Viennot) s'inscrivant parfois en contradiction avec les connaissances scientifiques*** [99] ***.*** *»*

4.5. Chevallard et Wozniak

Chevallard a remis en 1978 un rapport : « Notes pour la didactique de la statistique » dans lequel

---

[98] la statistique et les probabilités

[99] souligné par nous.

il pointe une difficulté de la transposition didactique :

> *« L'analyse statistique intervient toujours dans un champ de savoir structuré par certaines problématiques propres, - un territoire -, et la réduction didactique, en l'extirpant de ce territoire la prive de son sens, et ne lui laisse guère qu'un intérêt le plus général, et donc le plus pauvre aussi. Tel est l'effet de cette déterritorialisation. Celle-ci a des effets d'autant plus pernicieux qu'elle est mieux cachée : un nombre imposant d'ouvrages se présentent aujourd'hui comme des introductions à la statistique pour le psychologue, pour le géographe, etc. En fait, à peu de chose près, ces ouvrages disent la même chose. Le champ extra-statistique auquel ils se réfèrent chacun, n'est là que pour fournir des illustrations, et non des problèmes. »* Chevallard (1978), cité par Régnier.

Dans un article commun avec Chevallard (2003 b) puis dans sa thèse (2005), Wozniak pointe nettement, pour expliquer l'échec relatif de l'enseignement des probabilités et de la statistique, l'incapacité de l'enseignement secondaire des mathématiques à prendre en charge les *mathématiques mixtes* :

> *« Tout un ensemble de travaux (…) conduits dans le cadre de la théorie anthropologique du didactique, s'étaient articulés à une notion ancienne et quelque peu oubliée, celle de mathématiques* ***mixtes****, pour penser et approfondir la résistance de l'enseignement des mathématiques à la mixité objectale et au métissage épistémologique* [100] *.*
> *De ce point de vue, le renouvellement de l'étude de la statistique en seconde allait une fois de plus* [101] *mettre en scène l'antique refus, quasiment consubstantiel à notre enseignement des mathématiques, d'entretenir un commerce vivant avec des réalités tenues pour non mathématiques. »*

5. Une ébauche de synthèse de ces points de vue

Fischbein et Darmois convergent quand ils diagnostiquent une manière inadéquate d'aborder certains problèmes : déformation par un enseignement essentiellement déterministe selon Fischbein, méconnaissance d'une autre manière d'interpréter les faits observés que dans une vision déterministe, intégrant l'aléatoire dit Darmois.

Régnier essaye de définir un *esprit statistique*, à acquérir, donc à enseigner, qui serait caractéristique d'une démarche spécifique à la statistique .

Ces trois auteurs soulignent donc plutôt la spécificité d'une démarche.

---

[100] « sur ce thème, voir Chevallard (2001b) »

[101] Wozniak parle des programmes 2000.

Maury, Chevallard et Wozniak soulignent le rapport qu'entretiennent avec la réalité la statistique et les probabilités.

Maury pointe les connaissances spontanées de l'élève, acquises en dehors de l'institution scolaire, avec lesquelles l'enseignant devra composer, c'est peut-être l'une des difficultés de l'enseignement de ces disciplines.

Pour Wozniak, l'enjeu serait plutôt d'ordre culturel dans la tradition de l'enseignement des mathématiques en France, consistant en l'*épuration* de la statistique et des probabilités de ce qui devrait en faire des *mathématiques mixtes* : épuration du lien avec les objets réels qu'il s'agirait d'étudier. Les simulations seraient une occasion d'évacuer « en douce » ce lien, en permettant de faire semblant de traiter du réel.

Les deux points de vue ne paraissent pas s'opposer : évacuer le lien de ces disciplines avec la réalité, conduit à une *épuration* pour rester dans les *mathématiques pures*, mais évacue également la difficile prise en compte des connaissances spontanées des élèves, qui sont elles construites dans des situations extérieures à l'institution scolaire. Qui est premier dans cette affaire ?

Il semble de plus que ces domaines de la statistique et des probabilités, comme nous l'avons souligné dans la première partie pour la loi des grands nombres, sont des domaines qui sont particulièrement difficiles et sensibles pour l'enseignant <u>dans son propre rapport à ce savoir qu'il doit enseigner</u>, parce qu'il parle peu ou prou du hasard.

Par ailleurs, les domaines qui entretiennent un lien fort à la réalité ont disparu progressivement de l'enseignement des mathématiques en lycée : la cosmographie, la cinématique.
Toute une génération d'enseignants a enseigné des mathématiques épurées, comme dit Wozniak ; certains ayant été formés dans la période des *maths modernes*, n'en ont peut-être retenu qu'une image idéalisée des mathématiques, unifiées dans la théorie des ensembles, et ont peut-être transmis cette image aux futurs enseignants de mathématiques qui étaient leurs élèves.

Des enseignants actuellement en activité peuvent-ils facilement s'imaginer assurant un

enseignement qui parle de la réalité alors que toute leur formation scolaire, leur formation universitaire probablement, et peut-être aussi leur formation d'enseignant, les ont enfermés dans *les mathématiques pures* (ou *des mathématiques pures*) ?
N'y a-t-il pas là aussi une question sensible pour l'enseignant dans son propre rapport au savoir si ce savoir inclut des *mathématiques appliquées* ?

Ceci explique peut-être que certains enseignants n'acceptent pas la statistique et les probabilités comme filles des mathématiques.
La place de la statistique et des probabilités dans les mathématiques paraît en effet ne pas faire consensus dans le corps des enseignants du second degré : ces domaines ne seraient-ils pas extérieurs au mathématiques, les mathématiques n'y intervenant qu'en termes d'outils, comme pour les sciences physiques ? D'ailleurs un enseignement des statistiques et des probabilités a été longtemps assuré dans les universités de droit, de sciences économiques et de médecine plutôt que dans les études de mathématiques ; pourquoi ne pas continuer ?

Evidemment, ce point de vue pourrait s'appuyer sur ce qu'écrivait Bachelard (1938) à propos des obstacles que doit surmonter la connaissance empirique pour déboucher sur la connaissance selon l'*esprit scientifique* :

> *« On verra comment la **fausse rigueur** bloque la pensée, comment un premier système mathématique empêche parfois la compréhension d'un système nouveau. Nous nous bornerons d'ailleurs à des remarques assez élémentaires pour laisser à notre livre son aspect facile. D'ailleurs pour achever notre tâche dans cette direction, il nous faudrait étudier, du même point de vue critique, la formation de l'esprit mathématique. Nous avons réservé cette tâche pour un autre ouvrage. <u>A notre avis, cette division est possible parce que la croissance de l'esprit mathématique est bien différente de la croissance de l'esprit scientifique dans son effort pour comprendre les phénomènes physiques. En fait l'histoire des mathématiques est une merveille de régularité. Elle connaît des périodes d'arrêt. elle ne connaît pas des périodes d'erreurs. Aucune des thèses que nous soutenons dans ce livre ne vise donc la connaissance mathématique. Elles ne traitent que de la connaissance du monde objectif.</u>[102] »*

---

[102] passage souligné par nous.

Ce qui, si l'on suit Bachelard, signifierait que les mathématiques ne s'occupent pas de la connaissance du monde objectif (pas directement, puisque, tout de même, elles fournissent des outils pour les sciences dont l'objet est la connaissance du monde objectif).
La statistique et les probabilités s'occupant de produire des connaissances sur le monde objectif…
Ce débat paraît ici aussi vain que celui de la frontière entre les mathématiques et l'informatique.

L'important est de reconnaître qu'un adolescent arrive dans la classe de statistique ou de probabilités comme il arrive dans la classe de Physique : *« avec des connaissances empiriques déjà constituées : il s'agit alors, non pas **d'acquérir** une culture expérimentale, mais bien de **changer** de culture expérimentale, de renverser les obstacles déjà amoncelés par la vie quotidienne. »* Bachelard (1938), phrase déjà citée.

Ces obstacles ont été évoqués dans la première partie de cette thèse.
Les simulations informatiques peuvent être des outils pour dépasser certains d'entre eux.
C'est en partie l'objet des thèses de Zaki (1984), Bordier (1992) et de Coutinho (2001) par exemple.
C'est également l'objet de recherches plus ponctuelles telles que celle de Parzysc évoquée dans la première partie, celles d'Oriol et Régnier (Régnier (2009) par exemple).

Dans les « programmes 2000 », les simulations informatiques sont conçues d'abord comme des expériences aléatoires qu'on peut mettre en œuvre facilement, et répéter à satiété, afin d'introduire des notions : fluctuation d'échantillonnage, loi des grands nombres, modèle probabiliste.
Dans ce cadre, peu importe les expériences aléatoires d'origine qui sont simulées.
Il s'agit en fait de faire croire que les résultats observés lors de l'exécution des programmes informatiques sont les résultats d'une expérience aléatoire, pour pouvoir étudier *in vivo* cette expérience aléatoire là : inutile de revenir à l'expérience simulée.
Il est vrai que ces expériences aléatoires informatiques présentent bien des avantages de par la facilité de leur mise en œuvre, et de leur faible coût.

Mais nous partageons les craintes de Wozniak, si les simulations informatiques dans l'enseignement devaient n'être que cela : elles seraient effectivement un moyen commode d'éviter de se frotter à la réalité.

Dagnelie (2010) formule également des inquiétudes sur le risque que la facile répétition un grand nombre de fois d' une simulation informatique d'une expérience aléatoire, n'enferme les élèves dans la croyance qu'il est nécessaire de multiplier le échantillons pour faire des statistiques :

> *« Il y a en effet un risque réel de voir les élèves ou les étudiants considérer que le fait de disposer d'un nombre très important d'échantillons est la norme, alors qu'en pratique courante, l'utilisateur des méthodes d'inférence statistique ne dispose habituellement que d'un échantillon* ***unique****, pour la population ou pour chacune des populations étudiées. Le "nouveau néophyte", formé essentiellement par la simulation, risque de se trouver tout perdu quand il sera confronté en pratique à des ensembles de données constituées chaque fois d'un seul échantillon. »*

Nous pensons avec Dagnelie que cette « dérive » possible peut être évitée en plaçant les étudiants au moins une fois « *dans une situation "normale" d'application des méthodes d'inférence statistique* ». Nous ne développerons pas ici les situations mises en œuvre par Dagnelie avec ses étudiants. Il semblait utile de signaler ici une dérive semble-t-il rencontrée en Belgique avec des étudiants à l'université.

Nous allons essayer de montrer dans le chapitre qui suit que les simulations informatiques d'expériences aléatoires pourraient <u>aussi</u> être utilisées dans la formation en statistique et en probabilités pour ce qu'elles sont réellement dans une démarche scientifique : non pas un substitut facile d'emploi à de vraies expériences aléatoires permettant d'illustrer des notions d'un cours de statistique ou de probabilités, mais <u>une étape obligée dans un raisonnement pour étudier une situation de la réalité comportant une part d'aléatoire</u>, étape permettant

- non pas de prédire ce que dira le modèle probabiliste (cf. l'ordre des questions dans les sujets de l'épreuve expérimentale évoqués dans la première partie),
- ni de valider a posteriori ce qu'a dit le modèle probabiliste (cf. des essais d'enseignants, dans le cadre des IREM, relatant des travaux menés dans des classes),

mais d'apporter des connaissances sur la situation que l'on veut étudier, <u>que l'on n'a pas su trouver ailleurs</u>.

Chapitre 22

Modèles informatiques et théorie des probabilités dans un contexte de résolution de problème

Notre thèse est ici que les simulations ouvrent des perspectives au delà d'une simple propédeutique préparant à une *vraie* démarche probabiliste « plus tard » [103], et qu'elles ne peuvent être efficientes dans un enseignement de la statistique et des probabilités que si elles s'inscrivent résolument dans une problématique de mathématiques mixtes.

Cela impose que ne soit pas occultée la question des nombres pseudo-aléatoires, et qu'ils soient, de façon expérimentale, confrontés aux nombres aléatoires produits par de « vraies expériences » d'une part, et que d'autre part la notion d'*expériences équivalentes* soit travaillée *expérimentalement* avec les élèves : par exemple en comparant les résultats observés sur des expériences « vraies » et leurs simulations [104].

Nous avons essayé de montrer dans la deuxième partie de cette thèse que le lien entre une expérience aléatoire et sa simulation (informatique ou non) est leur *modèle probabiliste* commun, que ce modèle ait été calculé ou non, que la simulation soit obtenue par la technique de simulation de la loi de probabilité déjà calculée, ou que la simulation soit obtenue par la technique d'*imitation au plus près* de l'expérience aléatoire (l'équivalence entre l'expérience aléatoire et sa simulation étant alors *de principe*).

Il y a donc une mise en relation *théorique* entre la simulation d'une expérience aléatoire et son *modèle probabiliste*.
Théorique parce que le lien se fait, dans l'organisation praxéologique des deux tâches (modélisation et simulation de l'expérience aléatoire) du côte de la théorie, et très précisément de la théorie des probabilités : une *imitation au plus près* d'une expérience aléatoire n'est une

[103] Nous avons ici à nouveau en tête le constat de Lahanier-Reuter (1995) déjà évoqué au chapitre 2 paragraphe 4.

[104] Par exemple pour la somme des points amenés : lancers simultanés de deux dés, lancer deux fois de suite du dé, simulation du lancer de deux dés.

simulation de cette expérience que si elle a même *modèle probabiliste* que l'expérience.
Mais théorique aussi en ce sens que si l'on ne sait pas déterminer le *modèle probabiliste*, on ne pourra rendre « palpable » ce lien, il restera *théorique*, *virtuel*.

Les simulations, de natures diverses, sont un outil dans une démarche scientifique : on peut, pour de multiples raisons, décider, pour produire des connaissances sur une réalité, de la simuler.
Il n'est pas dans notre propos d'étudier les conditions de l'élaboration de ces simulations selon les domaines scientifiques et les diverses réalités qu'ils ont à étudier.
Il est vraisemblable qu'il y a toujours un jeu de va et vient entre la réalité et sa simulation, ce qui n'estpas le cas lorsque les simulations informatiques d'expériences aléatoires se substituent compètement à l'expérience initiale, sans retour sur cette expérience.

Il semble possible et souhaitable d'utiliser des simulations informatiques d'expériences aléatoires *es qualité*, c'est à dire dans une *vraie* démarche scientifique, pour résoudre un *vrai* problème.
Vrai problème, cela signifie une réalité crédible par les élèves amenant un problème susceptible de les motiver.
Vraie démarche scientifique signifiera alors appréhender le mieux possible l'incertitude liée au hasard, la solution retenue devant prendre en compte cette incertitude , si possible en la contrôlant, c'est à dire en la mesurant.
Cette démarche scientifique peut inclure une ou des simulations informatiques d'expériences aléatoires, qui sont alors présentes pour *produire des connaissances* <u>*aidant à résoudre le problème.*</u>

Il s'agirait donc de proposer, par exemple, des problèmes de prise de décision dans des situations incertaines car comportant des résultats aléatoires, problèmes tels que le « problème de M. Dupond », ou le problème des tablettes de chocolat évoqué dans notre deuxième partie.
Ce pourrait être aussi des problèmes de validation des hypothèses de modélisation choisies pour résoudre un problème donné.
Et enfin bien sûr, il demeure les problèmes qui permettent d'introduire des notions du modèle probabiliste, ou de remédier à des conceptions erronées : validation expérimentale de la loi des grands nombres, fluctuation d'échantillonnage, intervalle de fluctuation à un seuil donné, etc. : la simulation informatique peut être, de fait, l'expérience aléatoire étudiée, l'expérience

aléatoire simulée devenant alors subsidiaire, avec les risques soulignés par Wozniak. Il ne s'agit peut être plus alors de résoudre un problème, mais d'observer des résultats pour illustrer un théorème admis, ou remédier à une conception fausse.

Nous n'avons pas la prétention de faire une liste exhaustive de « types de problèmes » à résoudre qui permettraient de placer les élèves dans des situations qui pourraient, peut-être, être envisagées en termes de *situations didactiques* telles que définies par Brousseau (1998).
Nous nous contenterons de suggérer des connaissances que les simulations informatiques peuvent introduire grâce à certains « types de problèmes » que nous allons évoquer, en essayant de les présenter en termes de tâches, relevant de types de tâches, avec les praxéologies que nous pensons pouvoir leur associer.
Nous avons fait ce travail pour des problèmes de prise de décision avec un risque choisi (en 1.1.), et des problèmes de validation d'hypothèses de modélisation (en 1.2.).

Il n'est pas question ici de définir le scénario de *situations didactiques*.

1. Problèmes de prise de décision dans des situations dépendant du hasard

1.1. Problèmes de choix d'un paramètre

Nous avons développé assez longuement le problème de M. Dupond. Le problème des tablettes de chocolat relève aussi de ce type de problème.
Nous allons tenter de définir ces problèmes comme des tâches relevant d'un même type de tâche, pour essayer de proposer des techniques de résolution, et voir quelles technologies et théories les accompagnent.
La tâche consiste à décider du choix d'un nombre X pour satisfaire un ensemble E de contraintes, alors que le contexte comporte une part d'aléatoire.
Dans le cas du problème de M. Dupond, il s'agit de décider du nombre X de cases à mettre sur le plateau de jeu pour que la partie soit terminée à l'issue de 20 tours.
Dans le cas des tablettes de chocolat, il s'agit de déterminer le nombre X de tablettes de chocolat qu'il faudra acheter pour avoir la collection complète d'images.

choix de X			
	techniques τ	**technologies θ**	**théories Θ**
(1)	• choisir le risque α accepté	• définition d'un indicateur mesurant le risque que les contraintes ne soient pas toutes satisfaites en faisant le choix x_0 pour X	
(2)	• « extraire » du problème une expérience aléatoire dans laquelle les contraintes à respecter E sont prises en compte, et où apparaissent une ou plusieurs variables aléatoires sur lesquelles sera fondé le choix de X.	•	
(3)	• simuler cette expérience un grand nombre de fois (n fois), et récupérer les séries des n valeurs observées pour chacune des variables aléatoires.	• faire une imitation au plus près de l'expérience. *voir le chapitre 19*	• *voir le chapitre 20*
(4)	• Faire une étude statistique de ces séries de n valeurs,	• indicateurs statistiques connus / les variables semblent-elles suivrent une loi connue ?	• statistique descriptive ; variables gaussiennes
(5)	• et décider d'une valeur x_0 pour X : le choix se fait de telle sorte que lors des n simulations, avec cette valeur de X, la proportion des simulations pour lesquelles les contraintes ont été satisfaites est au moins égale à $1 - \alpha$.	• si c'est possible, vérifier directement que pour les n simulations, avec <u>cette valeur</u> x_0 pour X, les contraintes n'ont pas été satisfaites pour une proportion au plus α, sinon modifier la simulation en prenant en compte cette valeur x_0 et la répéter n fois en faisant calculer pour chaque essai une variable « succès ».	
(6)	• envisager les effets de la fluctuation d'échantillonnage .		• fluctuation d'échantillon-nage

Les 6 étapes ici définies appellent quelques commentaires

(1) : il faut d'abord s'assurer que l'on peut, selon le choix de X, faire varier le risque de ne pas

satisfaire l'ensemble E des contraintes. Dans le problème de M. Dupond, si l'on met la case d'arrivée sur la case départ, ce risque vaut zéro. Si l'on met plus de 60 cases, le risque est de 100 %, puisqu'à chaque coup, un joueur avance au plus de 3 cases.

Cela étant établi, le niveau de risque à choisir dépend de différents paramètres, objectifs et subjectifs (goût du risque du décideur).

(2) En lycée, il semble que l'on doit pouvoir trouver des problèmes, tel les deux problèmes ici évoqués, pour lesquels l'expérience aléatoire s'impose pratiquement d'elle-même :

- Exp1 : on fait une partie de 20 tours avec trois joueur pour le problème Dupond,
- Exp2 : on achète des tablettes tant que la collection des 6 images n'est pas complète,

Dans Exp1, on peut dégager 4 variables aléatoires (X_i = le nombre de cases avancées par le ième joueur, et Y = max (X_1, X_2, X_3).

Dans Exp2, on peut dégager la variable N = le nombre de tablettes à partir duquel la collection a été complète.

(3) Nous nous plaçons ici dans le cas où le choix a été fait de faire une modélisation informatique de l'expérience, donc la simulation sera une *imitation au plus près* telle que nous l'avons définie dans la deuxième partie.

(4) Il y a intérêt bien sûr à collecter les séries numériques dans un logiciel permettant de calculer les indicateurs statistiques souhaités en utilisant directement les fonctionnalités statistiques du logiciel, ou une bibliothèque de logiciel.

Naturellement, si l'on remarque qu'on peut faire l'hypothèse par exemple que l'une des variables est gaussienne, on peut utiliser les résultats, s'ils sont connus, sur les variables gaussiennes.

(6) Il est possible (nous l'avons vu avec Exp1) que si n n'est pas assez grand, la fluctuation d'échantillonnage puisse remettre en cause le choix x_0.

Pour ce type de problème, les connaissances de la théorie des probabilités sont très limitées ; l'objectif de ces problèmes pourrait être surtout de familiariser avec les grandes questions des domaines de la statistique et des probabilités que sont l'incertitude, le risque, la fluctuation d'échantillonnage, et amener à traiter ces problèmes avec une démarche nouvelle (l'esprit probabiliste ?).

1.2. Problèmes de tests d'hypothèses

Il peut s'agir, par exemple, de valider l'hypothèse que pour une expérience aléatoire donnée,

la loi de probabilité est équirépartie, ou plus précisément que le modèle choisi pour l'expérience, modèle calculé sous l'hypothèse que la loi est équirépartie, est acceptable.
Ces problèmes sont dans les programmes 2000 (en terminale), mais semblent disparaître des programmes 2011, ce que l'on peut regretter, car ces problèmes permettent d'introduire concrètement assez tôt (éventuellement avant la terminale) les risques de première espèce.

Roditi (2009) a mis en œuvre une démarche analogue avec des étudiants en sciences humaines pour introduire le test du χ^2 (la feuille de calcul est conçue par l'enseignant, mais l'étudiant à accès à toutes les formules, ce qui doit lui permettre de comprendre les étapes du raisonnement).

Il s'est exprimé, dans la noosphère, des points de vue hostiles au maintien dans les programmes de ce paragraphe des « programmes 2000 » concernant le test de l'hypothèse d'équiprobabilité s'appuyant sur des simulations, et il est vrai que les exercices proposés au baccalauréat sur ce thème rivalisaient souvent dans le ridicule tellement ils étaient peu crédibles et caricaturaux de situations « pseudo-réelles ».

La problématique paraît pourtant correspondre à une demande des élèves (c'est ce que nous avons eu l'occasion d'observer), lorsqu'on se donne la peine de creuser un petit peu la question du choix des hypothèses de modélisation quand on veut déterminer une loi de probabilité, ne serait-ce que celle d'un simple dé cubique : que faire lorsqu'on soupçonne que ce dé est mal équilibré ? Des élèves proposent spontanément de comparer un grand nombre de lancers de ce dé avec le même nombre de lancer d'un dé équilibré.

Ces problèmes de tests d'hypothèses, abordés avec des simulations informatiques, semblent propices pour introduire la notion de risque de première espèce, et une réflexion sur le fait que lorsqu'on accepte une hypothèse H_0 au risque α, $1 - \alpha$ ne donne pas la probabilité d'avoir raison d'accepter H_0 : α mesure, lorsqu'on la rejette, la probabilité d'avoir tort de la rejeter. Il est possible de mener un travail de logique avec les élèves sur cette question difficile.

valider une hypothèse		
techniques τ	**technologies θ**	**théories Θ**
• recueillir des données expérimentales sur un échantillon E de taille n de l'expérience aléatoire : si cette expérience a k issues, calcul de la distribution des fréquences des issues $(f_1,\ldots, f_k)$		• risques de première et deuxième espèce
• définir l'hypothèse H_0		
• choisir le risque α accepté	• il s'agit de définir le risque α de rejeter <u>à tort</u> H_0.	
• définir un indicateur d_{obs} qui mesure la distance entre la distribution $(f_1,\ldots, f_k)$ de E à la distribution $(\frac{1}{k} ; \ldots ; \frac{1}{k})$ équirépartie	• d peut être la somme des valeurs absolues des écarts, ou la somme des carrés des écarts, ou le plus grand écart, ou l'un de ces nombres multiplié par n	• la multiplication par n rend l'indicateur peu sensible à n
• simuler N fois un échantillon de taille n d'une loi équirépartie à k issues, et calculer pour chacun des échantillons l'indicateur d correspondant	• N doit être grand : en ce cas, d'après la loi des grands nombres, la proportion des éléments de cet échantillon qui dépassent une valeur a > 0 donnée est voisine de la probabilité de l'événement Δ : « dans un échantillon de taille n d'une loi équirépartie à k issues, d dépasse a »	• loi des grands nombres
• Calculer le (100 - α) centile D de la série statistique des N valeurs des échantillons simulés. • si d_{obs} >D, rejeter H_0 au risque α de se tromper, sinon accepter H_0.	• si H_0 est valide, l'échantillon E des données sur l'expérience de départ est aussi un échantillon de la loi équirépartie ; la probabilité que le d de cet échantillon dépasse D est donc environ α, donc lorsqu'on rejette H_0 parce que d_{obs} > D, on prend le risque α de le rejeter à tort.	

2. Problèmes permettant d'illustrer le cours de probabilité

Les simulations informatiques d'expériences aléatoires peuvent illustrer certaines notions du cours de probabilité.

On peut trouver beaucoup de productions, particulièrement du côté des IREM, sur ce point.

Nous nous limiterons à tenter de définir quelques types de problèmes, mais sans essayer de préciser une praxéologie comme nous l'avons fait dans le paragraphe précédent :

- problèmes illustrant la fluctuation d'échantillonnage,
- problèmes illustrant la loi des grands nombres,
- problèmes illustrant le lien entre la valeur moyenne observée sur un échantillon de grande taille d'une variable aléatoire et l'espérance mathématique de cette variable,
- problèmes illustrant la plus ou moins grande dispersion des valeurs prises par des variables aléatoires ayant même espérance autour de cette espérance commune, selon la valeur des écarts- types des variables,
- problèmes illustrant les notions d'intervalle de confiance et niveau de confiances pour des variables, gaussiennes, ou non (programmes de seconde 2010).

CHAPITRE 23

Qui n'est pas tout à fait une conclusion

Nous sommes consciente d'avoir laissé en suspens un certain nombre de questions, en particulier, la question suivante : qu'est-ce qui assure qu'une *imitation au plus près* de l'expérience aléatoire est bien une simulation de l'expérience aléatoire, c'est à dire une autre expérience aléatoire **qui a même modèle que l'expérience initiale**, lorsque nous revendiquons que cette imitation, précisément, ne se construit pas à partir du *modèle probabiliste* de l'expérience aléatoire ?

Nous allons d'une part essayer de montrer qu'un certain nombre d'auteurs accordent peut-être trop d'importance à cette question, et d'autre part proposer quelques pistes pour y répondre.

1. Les simulations informatiques, « un cercle vicieux didactique » ?

A propos des programmes 2000, certains auteurs soulignent ce qu'ils considèrent comme une cercle vicieux :

> *« La méthode pédagogique préconisée fait une large place à la simulation. Cette nouvelle approche pose de nouveaux problèmes didactiques liés aux difficultés propres à l'activité de modélisation et à l'ambiguïté du statut de la simulation. (...)*
>
> *Au moyen d'un habillage éventuellement évocateur d'une situation aléatoire concrète* [105] *, l'ordinateur est donc implicitement considéré comme un générateur de hasard qui, correctement programmé, fournit des résultats parfaitement représentatifs des issues que l'expérience concrète évoquée fournirait. Cette assimilation ne va pas de soi pour tous les élèves (cf. biais psychologiques évoqués plus haut). L'ordinateur donne* ***au hasard*** *des chiffres de 1 à 6 équirépartis, mais* ***ce n'est pas un vrai dé****, même non pipé. L'équivalence pour être admise suppose implicitement l'idée que la probabilité de sortie de l'as sur le dé est égale à celle de sortie du 1 de l'ordinateur, c'est à dire que le concept de probabilité soit installé. L'utilisation de l'ordinateur pour introduire la notion de*

[105] Ceci va dans le sens du constat que nous avons fait dans le précédent chapitre : les expériences aléatoires que sont les simulations informatiques sont étudiées pour elles-mêmes, il n'y a pas de retour sur les expériences simulées, qui servent de prétextes.

probabilité, comme représentant une expérience concrète, repose donc sur un cercle vicieux didactique, ou bien sur un contrat didactique préalable impossible à expliciter : c'est bien le hasard du dé qui est représenté par l'ordinateur. (...) Rigoureusement, sans modèle théorique [106]*(puisque la notion de probabilité est exclue) il n'est pas question de simulation. Le terme semble abusif en seconde quand il s'agit seulement de* ***représenter*** *le comportement des issues d'une expérience concrète. Mais on peut penser qu'il y a un modèle sous-jacent, dès lors que la production des chiffres aléatoires par l'ordinateur n'est pas exploitée n'importe comment. Dans les situations simples, comme le lancer d'un dé ou tout autre générateur standard, l'équiprobabilité est implicitement admise et associée spontanément à la loi uniforme discrète censée gouverner la production des chiffres aléatoires. Le terme de* **simulation** *peut donc être justifié en référence à un savoir minimal probabiliste, mais pas explicitement à des élèves de seconde. (...) Ce qui peut garantir cette équivalence* [entre une expérience et sa simulation] *c'est que les deux expériences relèvent du même modèle probabiliste, indisponible à ce moment là. Comment interpréter alors les fluctuations d'échantillonnage observées dans la répétition de l'expérience simulée comme inhérentes au caractère aléatoire de l'expérience étudiée ? Sans réponses à ces questions, le professeur ne peut que s'en tenir à un écran de fumée supposant que les élèves ne se les poseront pas (ce qui n'est pas une hypothèse absurde).»* Girard et Henry (2005).

Ces questions appellent plusieurs remarques.

<u>1.1. L'équivalence pour être admise passe-t-elle par le concept de probabilité ?</u>

« L'équivalence pour être admise suppose implicitement l'idée que la probabilité de sortie de l'as sur le dé est égale à celle de sortie du 1 de l'ordinateur, c'est à dire que le concept de probabilité soit installé. » disent Girard et Henry. Cela est vrai au niveau théorique du professeur par exemple, s'il a comme définition qu'une simulation est une expérience aléatoire ayant même modèle probabiliste que l'expérience simulée.

Ce n'est peut-être pas la définition d'une simulation qu'il est souhaitable – au moins dans un premier temps – de donner à un élève de seconde, ou même en début de première (et a

[106] Girard et Henry parlent des programmes de seconde.

fortiori de collège).
La définition pourrait être plutôt, dans des situations concrètes, par exemple de jeux, liée au fait de remplacer pour **un** joueur une expérience aléatoire par une autre en conservant l'équité du jeu, ou en faisant comparer des indicateurs statistiques [107] associés à n épreuves (n plus ou moins grand) d'expériences que l'on voudrait voir qualifier par les élèves d'*expériences équivalentes*.

Nous faisons faire en classe de seconde ou de première ces calculs par nos élèves avec les trois expériences :
1. somme des points obtenus lors du lancer de deux dés cubiques
2. somme des points obtenus en lançant deux fois de suite le même dé
3. simulation sur tableur de la somme des points obtenus en lançant deux dés.

La troisième expérience nous permet d'amener une réflexion
• d'une part sur les hypothèses de modélisation : les dés sont équilibrés,
• d'autre part sur la différence entre une imitation des seules issues (2, 3, …., 12), et une imitation de ces issues avec leurs « chances ». Il y a toujours eu dans nos classes au moins une « simulation » par une loi équirépartie sur les issues, ce qui est l'occasion d'un débat, précisément sur ce que nous devons entendre par *simulation*.
La situation a toujours été intéressante :
• soit que les élèves remarquent d'eux-mêmes (quand ils choisissent de présenter une loi équirépartie sur les issues) qu'il y a quelque chose qui ne va pas, car le douze *devrait* sortir moins souvent que le sept, et c'est souvent l'occasion pour l'élève de commencer à essayer de représenter mathématiquement *les issues physiques*, mais pas toujours en distinguant (2 ; 5) et (5 ; 2),
• soit que les élèves remarquent que les distributions des fréquences pour cette « simulation » sont sensiblement différentes de celles obtenues dans les deux premières expériences, au delà d'une fluctuation d'échantillonnage, mais sans remettre en cause le statut de *simulations* à ce qu'ils ont créé.

<u>1.1. Peut-on faire une simulation sans le savoir ?</u>

Comme Monsieur Jourdain faisait de la prose…

---

[107] avec au moins un indicateur pertinent de dispersion ; on peut aussi présenter des diagrammes de fréquences.

Deux expériences aléatoires, pour être équivalentes, doivent-elles avoir été construites comme telles ?
Ou encore : lancer un dé non pipé et tirer « au hasard » un papier parmi 6 papiers bien mélangés numérotés de 1 à 6, ne sont-elles des expériences équivalentes qu'à partir du moment où quelqu'un les déclare telles parce qu'elles ont même modèle probabiliste ?

Du point de vue des élèves [108] , il semble important qu'ils adhèrent plutôt à une équivalence entre l'expérience et sa simulation de façon expérimentale, en travaillant sur quelques exemples simples.
C'est cette adhésion, à partir de quelques expériences simples et de leurs simulations, qui peut permettre de constater, sur ces exemples effectivement manipulés, qu'une expérience et sa simulation ont même loi de probabilité [109].
Les expériences élémentaires mises en jeu au lycée sont des lancers de dés, de pièce, des tirages d'urne, pour lesquelles il est possible (et souhaitable) de faire vérifier l'équivalence de la simulation informatique avec l'expérience élémentaire, au sens par exemple : on peut remplacer l'une par l'autre pour l'un des joueurs et le jeu reste équitable.
Cette équité doit être constatée, nous l'avons déjà dit, pas seulement en comparant des valeurs moyennes, et des dispersions sur de grands échantillons, mais en comparant les fluctuations d'échantillonnage pour des échantillons de tailles comparables.

Pour reprendre la question posée par Girard et Henry, il semble qu'alors les fluctuations d'échantillonnages observées dans les simulations d'expériences plus complexes peuvent être considérées par les élèves comme représentatives des fluctuations de l'expérience simulée, y compris lorsque les élèves ont compris que le « hasard » d'une calculatrice est calculé.

2. De l'imitation à la simulation

Cela étant, la question qui nous est posée est plutôt : comment concevoir, mettre en place, une expérience aléatoire équivalente à une expérience aléatoire donnée, sans passer par le modèle probabiliste, suivie de la question : et comment s'assurer que cette expérience est bien

[108] Tout au moins dans un premier temps.

[109] Il est peut-être plus simple de déterminer la loi de probabilité d'une simulation non informatique …
Par exemple deux dés simultanément simulés par un dé lancé deux fois, ou deux tirages de papier dans deux chapeaux contenant des papiers numérotés de 1 à 6.

équivalente à la première ?
Nous avons pour l'instant répondu à ces questions en proposant de *faire une imitation au plus proche*, de *faire un cahier des charges* qui repère les moments de l'expérience où « se joue l'aléatoire » en les décomposant en expériences aléatoires élémentaires, et qui repère les différentes actions en jeu (actions aux résultats aléatoires ou déterministes), leurs liens et leurs enchaînements chronologiques ; nous avons souligné l'étape de pré-modélisation où doivent se faire les hypothèses de modélisation sur les expériences aléatoires élémentaires.
Alors, avons-nous dit, l'équivalence sera *de principe*.
Nous allons tenter de préciser cela.

2.1. Le cahier des charges

Cette expression a cours en informatique (cf. chapitre 8, paragraphe 2), par exemple lorsqu'un non informaticien (NINF) passe une commande à un informaticien (INF) pour résoudre un problème donné : NINF doit spécifier le plus précisément possible ce qu'il attend du programme : quelles seront les données, sous quelles formes, quels devront être les résultats, sous quels formes, quels paramètres devront pouvoir être modifiés par NINF, etc.
Par exemple, NINF, professeur, peut commander à INF un carnet de notes matérialisé par une feuille de calcul sur tableur, où NINF pourra entrer jusqu'à 40 noms (type chaînes de caractères), avec en regard jusqu'à 6 notes sur 20, et pourra choisir les coefficients à affecter à ces notes. Il faudra qu'il récupère dans une autre feuille du même fichier la liste avec en regard des noms la moyenne coefficientée, et dans un cadre à part de cette feuille les indicateurs statistiques qu'il choisira parmi une liste [110].

Nous allons présenter le cahier des charges pour le « problème de Monsieur Dupond » . Ce cahier des charges est défini dans l'encadré ci-dessous.

Il faut simuler ceci :
• on joue 20 tours
• pendant lesquels chaque joueur (il y en a 3) :
lance deux dés tétraédriques et calcule l'écart absolu entre les points amenés, avance d'autant de cases sur le plateau que cet écart
• on regarde à l'issue de ces 20 tours de combien de cases a avancé celui des trois joueurs qui a été le plus loin.

[110] Ceci est l'ébauche d'un « cahier de notes interactif »…

A y regarder de près, on réalise qu'on pourrait transformer ce texte en :

Il faut modéliser ceci :
• on joue 20 tours
• pendant lesquels chaque joueur (il y en a 3) :
lance deux dés tétraédriques et calcule l'écart absolu entre les points amenés, avance d'autant de cases sur le plateau que cet écart
• on regarde à l'issue de ces 20 tours de combien de cases a avancé celui des trois joueurs qui a été le plus loin.

Ce qui, compris dans une perspective de *modélisation probabiliste* amène la solution suivante déjà évoquée :

Définissons, pour un tour et pour un joueur, la variable aléatoire X égale au nombre de cases avancées par un joueur.
Nous devons faire des hypothèses de modélisation sur les dés : supposons que les dés sont équilibrés. La loi de probabilité associée à chacun des dés est alors équirépartie, et un calcul simple permet de présenter la loi de probabilité de X dans le tableau suivant :

x_i	*0*	*1*	*2*	*3*
$p(X=x_i)$	*4/16*	*6/16*	*4/16*	*2/16*

L'espérance de X est $E(X) = 1,25$.
Pour un joueur A qui joue 20 tours, on considère $X_A = \sum X_i$ où X_i est la variable aléatoire égale au nombre de cases avancées par ce joueur au i-ème tour. Pour tout i, $E(X_i) = E(X) = 1,25$, donc par linéarité de l'espérance, $E(X_A) = 20 \times 1,25 = 25$.
Pour trois joueurs A, B, C on définit trois variables aléatoires X_A, X_B et X_C ayant toutes même espérance 25.
Le nombre de cases avancées, à l'issue de 20 tours, par celui des trois joueurs qui a été le plus loin est la variables aléatoire $Y = \max(X_A, X_B, X_C)$.
Calculons l'espérance de Y.

Et là on constate que Y prend ses valeurs dans l'ensemble {0 ; 1 ; …. ; 60}, et que, s'il n'est pas impossible de calculer l'espérance de Y, ce calcul promet d'être fastidieux.

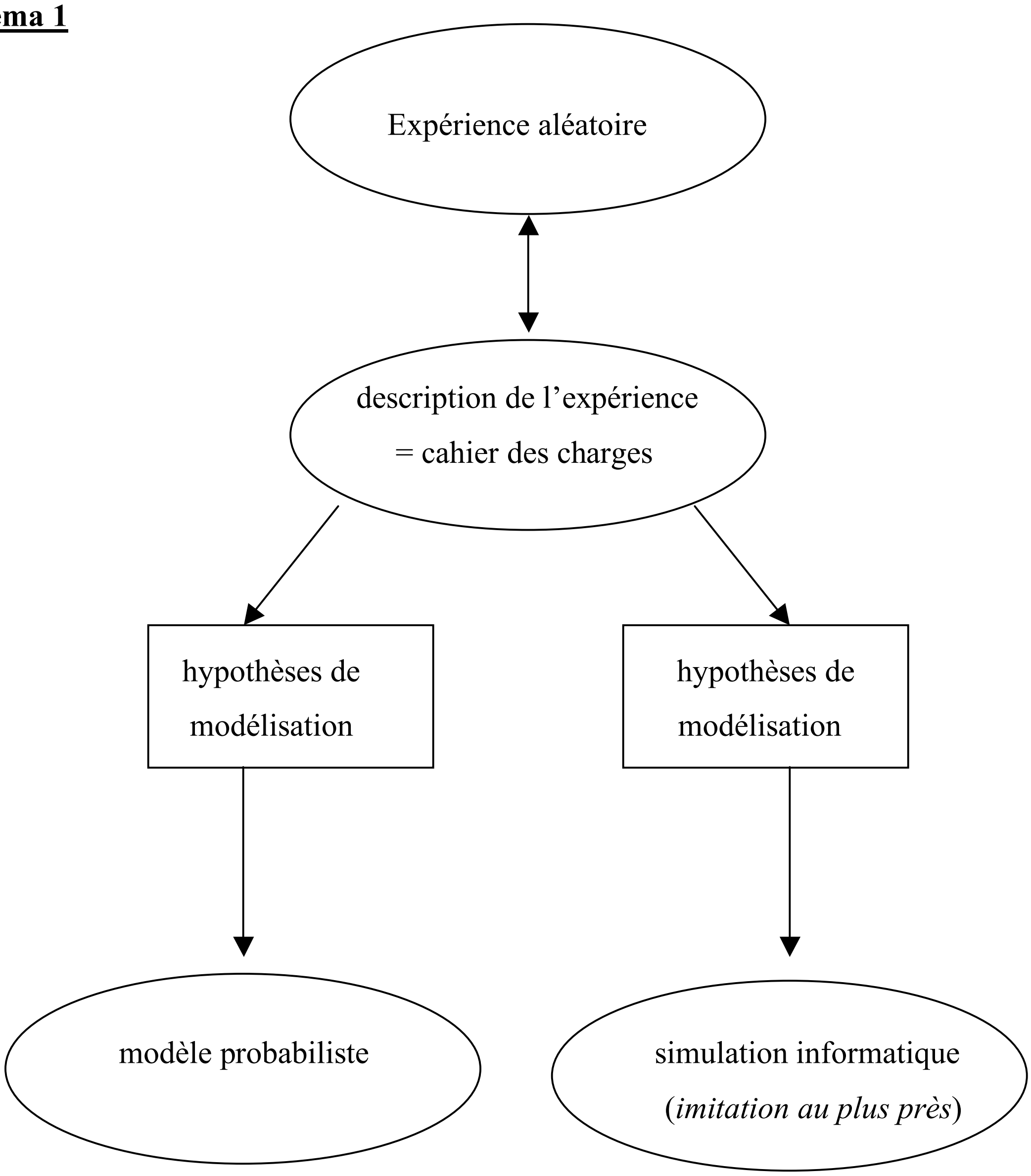
Nous avons voulu montrer sur cet exemple que le cahier des charges derrière lequel nous avons semblé nous réfugier, correspond ici à la description de l'expérience aléatoire telle qu'on peut l'attendre lorsqu'on en entreprend une *modélisation probabiliste*.
Il semble que ceci se vérifie pour toutes les expériences aléatoires avec un nombre fini d'issues, ou un ensemble d'issues dénombrable, qu'il est raisonnable d'étudier en lycée pour tenter d'en construire un modèle probabiliste dans le cadre des « programmes 2000 ».
Nous ne nous aventurerons pas à sortir du cadre du lycée, ni à généraliser, pour le lycée, ce constat pour des expériences qui relèveraient de lois à densités continues.

Revenons alors au cahier des charges tel que défini plus haut pour le « problème de M. Dupond ». Une fois fait le choix d'hypothèses de modélisation sur les deux dés, et le style de programmation visé, la description d'algorithme est complètement déterminée par ce cahier des charges, et partant le programme.
C'est en ce sens que nous avons parlé de simulation *par une imitation au plus près*, au plus près de la description de l'expérience aléatoire, cette *imitation au plus près* étant pratiquement complètement déterminée dès qu'est défini le cahier des charges et que sont définies les hypothèses de modélisation, de même que le modèle probabiliste est complètement déterminé dès qu'est défini la description / cahier des charges de l'expérience et que sont définies les hypothèses de modélisation.

2.1. L'*imitation au plus près* a même modèle que l'expérience simulée

Ce qui précède amène le constat suivant : quand nous disons que la simulation par *imitation au plus près* est plus qu'une imitation des issues, mais une expérience aléatoire ayant même modèle que l'expérience aléatoire simulée, alors qu'elle n'est pas construite à partir de ce modèle, cette affirmation est mieux qu'une pétition de principe.
L'expérience initiale et sa simulation par *une imitation au plus près* ont même modèle **car elles répondent au même cahier des charges, avec les mêmes hypothèses de modélisation**.
Nous avons employé le terme cahier des charges car c'est le terme adéquat lorsqu'il s'agit de définir les fonctionnalités du programme demandé à l'informaticien. Ce terme peut sembler déplacé lorsqu'il s'agit tout simplement de décrire une expérience aléatoire pour laquelle on veut définir un modèle probabiliste.
On peut faire le schéma suivant :

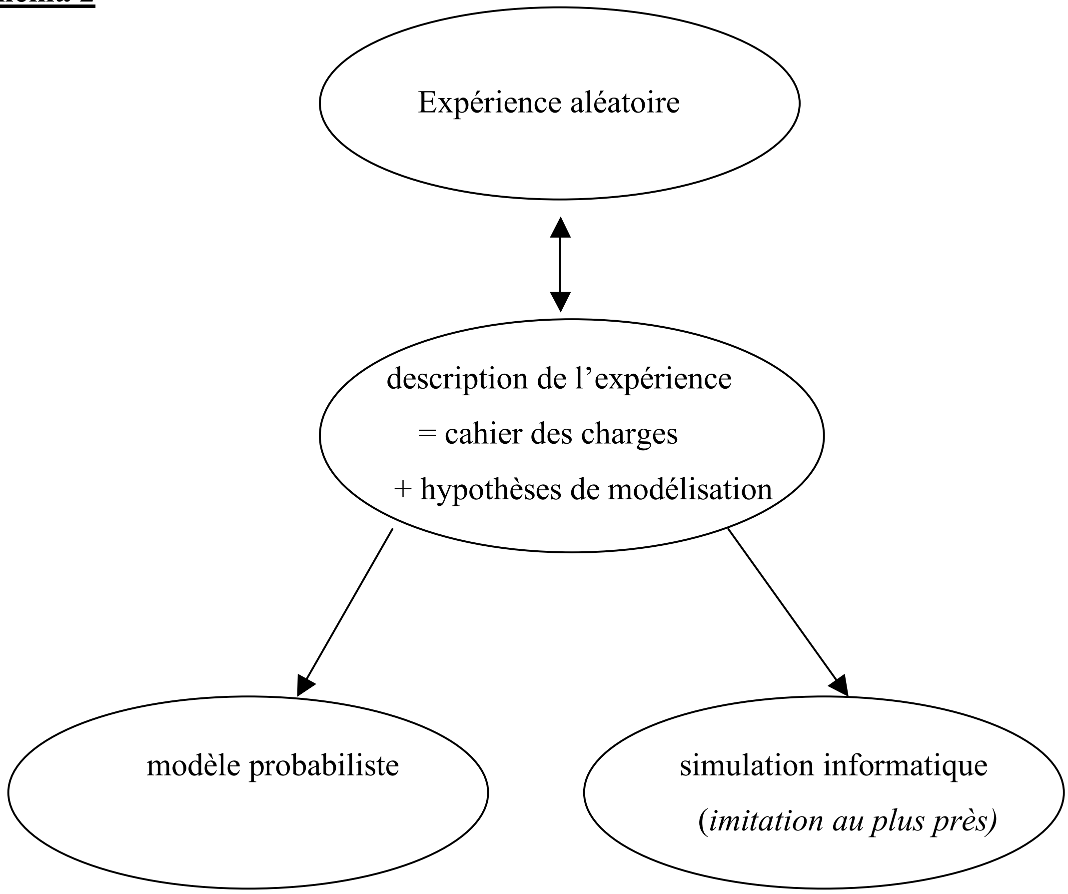
schéma 1



Nous avons représenté le lien entre l'expérience aléatoire et sa description par une double flèche : il y a une correspondance très étroite, que nous aurions envie de qualifier d'isomorphisme entre les deux, même si le terme est impropre.
La description peut précéder l'expérience au sens où elle peut la définir : imaginons un jeu en en définissant les règles, puis jouons ; c'est le cas dans le problème de Monsieur Dupond.
Il se peut que l'on n'ait même pas besoin de réaliser l'expérience.

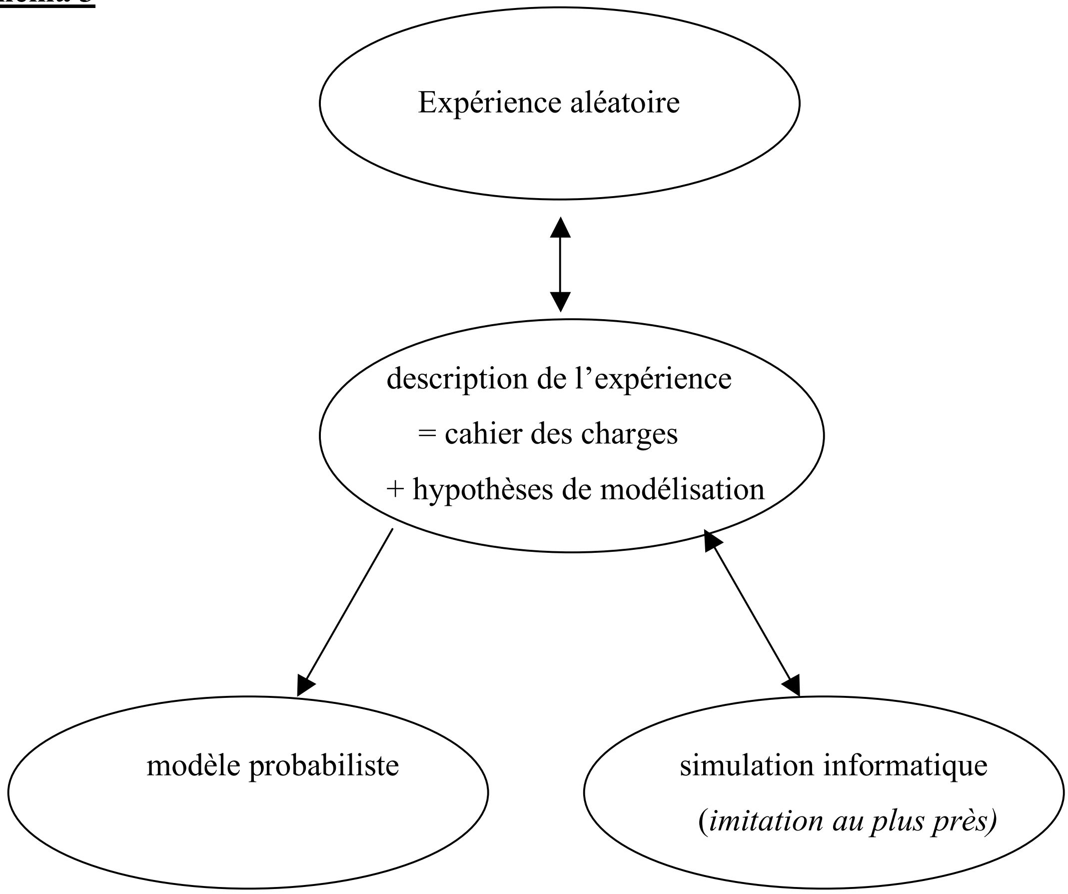
Par ailleurs, pour que la simulation soit une expérience aléatoire ayant pour modèle le modèle probabiliste en regard sur notre schéma, il faut que les hypothèses de modélisation soient les mêmes, d'où le schéma :

schéma 2



Mais l'imitation au plus près, qui est elle-même une expérience aléatoire, peut être décrite : c'est la description de l'expérience qui est au dessus : la simulation et l'expérience simulée ont même description et mêmes hypothèses de modélisation, d'où le schéma suivant, avec une double flèche entre la description munie des hypothèses de modélisation, et la simulation *imitation au plus près* :

schéma 3



La description de l'expérience <u>et les hypothèses de modélisation</u> menant à un seul modèle probabiliste, l'expérience simulée et sa simulation informatique par une *imitation au plus près* auront bien même modèle probabiliste.

Nous sommes consciente que nous ne pouvons affirmer que ce raisonnement vaut toujours et partout car il repose sur plusieurs hypothèses :

• l'identité entre la description de l'expérience et le cahier des charges de la simulation de cette expérience d'une part,

• le lien fort quasi d'isomorphisme entre une expérience et sa description d'autre part.

• la complète détermination de la description d'algorithme dès lors que le cahier des charges, les hypothèses de modélisation, <u>et le style de programmation sont définis</u> d'autre part.

Nous voulons souligner par ailleurs la pertinence de la double flèche entre l'*imitation au plus près* et la description/cahier des charges, ce qui indirectement valide notre identification entre la description et le cahier des charges : ce chemin de retour de la simulation vers le cahier des charges n'est pas qu'une commodité pour notre raisonnement ; il arrive que devant un programme informatique, on ait à définir ce que calcule le programme ; et si ce programme contient au moins un appel à une fonction aléatoire, définir ce que calcule le programme revient à décrire l'expérience aléatoire que génère son exécution. C'est ce que nous avons fait avec la simulation du « jeu du lièvre et de la tortue » dans le dernier paragraphe du chapitre 6.

Ce travail « du programme vers ce qu'il calcule » ou « d'une description d'algorithme vers ce que cet algorithme calcule » est un exercice préconisé dans la partie algorithmique - programmation des programmes 2010.

3. Puisqu'il faut conclure…

L'objet de cette thèse n'était pas de présenter une nouvelle voie pour enseigner les probabilités. Nous avons voulu voir à quoi pouvaient servir, dans l'enseignement en lycée de cette partie des mathématiques qui s'appelle « probabilités », des simulations informatiques d'expériences aléatoires.

Cela nous a amené tout naturellement d'abord à essayer de comprendre ce qu'est une simulation d'une expérience aléatoire, et plus particulièrement ses liens avec l'expérience simulée et le modèle probabiliste de cette expérience.

Nous avons compris que dès lors que les élèves, non seulement sont capables de programmer eux-mêmes un ordinateur, mais que c'est une capacité que l'institution scolaire a pour mission de leur faire acquérir, ces liens ont changé, et une simulation n'est plus la programmation d'un tirage d'urne qui aurait même loi de probabilité que l'expérience à simuler.

C'est en ce sens que nous avons pu écrire qu'il y avait changement de paradigme : la modélisation probabiliste de l'expérience n'est pas, n'est plus une étape indispensable à sa simulation informatique.

Il nous est apparu alors que si c'est l'élève qui doit créer la simulation, qu'il s'agisse d'un exercice *en soi* que propose l'institution, ou – et c'est notre point de vue – parce cette activité peut être efficace dans certains processus d'acquisition de connaissances en probabilité, il convient de réexaminer la tâche « simulation d'une expérience aléatoire » pour constater qu'elle relève **aussi** d'activités de type algorithmiques.

Nous pensons qu'il reste beaucoup de voies à explorer sur ces activités algorithmiques en lycée, et bien au-delà des simulations informatiques d'expériences aléatoires.

Nous avons voulu résolument centrer notre travail sur la didactique des mathématiques, et l'enseignement des probabilités en lycée. Nous nous sommes donc limitée à évoquer succinctement quelques problèmes didactiques concernant l'intervention de l'algorithmique dans la tâche de simulation.

Il nous est apparu que les « programmes 2000 » tendaient à enfermer les simulations d'expériences aléatoires dans le statut de substitut commode à de vraies expériences aléatoires, permettant d'illustrer la notion d'espace probabilisé et la loi des grands nombres, cela pour ce qui concerne leur intégration dans une progression didactique.

Quant à la présence de simulations informatiques dans les *épreuves pratiques expérimentales* qui ont eu lieu pendant quelques années, elle ne peut que laisser perplexe : les simulations précédaient toujours la loi de probabilité de l'expérience aléatoire (en contradiction avec la définition d'une simulation donnée par les programmes et les documents d'accompagnement), et ne semblaient avoir d'autre utilité que de justifier que l'exercice soit posé lors d'une épreuve « pratique » devant un ordinateur, puisqu'il n'était pas demandé d'établir un lien entre les résultats observés sur la simulation et le modèle probabiliste de la question suivante ; a fortiori il n'était pas demandé de commenter ce lien.

L'implication des élèves pour le « problème de Monsieur Dupond », constatée pendant la partie expérimentale de notre thèse, et la qualité inattendue des réponses de certains élèves, nous a amenée à essayer d'aller au delà d'une certaine frilosité de l'institution concernant ce qu'on peut envisager de faire avec une classe.
Divers résultats de cette expérimentation sur un type de tâche inédit pour ces élèves (et en temps très limité) : une prise de décision dans une situation d'incertitude, nous ont amenée à essayer de voir si ce type de tâche pouvait être intégré à une progression qui prendrait peut-être mieux en compte la spécificité de ce champ de probabilités, que cette spécificité s'appelle *esprit probabiliste* ou autrement.

Nous n'avons absolument pas la prétention d'avoir trouvé une « voie » idéale pour l'enseignement des probabilités. Il nous semble que la résolution d'un problème tel que « le problème de Monsieur Dupond » dont la solution s'appuie sur un *modèle informatique* de l'expérience aléatoire est une voie parmi d'autre pour introduire des notions de la théorie des probabilités, théorie des probabilités dont l'acquisition reste l'objectif de l'enseignement, ne serait-ce que parce que, un certain niveau étant atteint, cette théorie permet de se dispenser des simulations informatiques pour prendre les décisions : nous pensons par exemple aux tests, aux intervalles de confiance et niveaux de confiance tels qu'on peut les calculer dans le cadre de la théorie des probabilités.

Mais il nous semble indispensable de ré-évaluer les simulations dans l'enseignement des probabilités et de les concevoir résolument comme bien autre chose qu'un substitut à d'inaccessibles « vraies » expériences aléatoires.

BIBLIOGRAPHIE

Alarcon J. (1982). *L'appréhension des situations probabilistes chez des élèves de 12-14 ans : résultats d'une enquête proposée à des élèves de 4ème et 5ème*. Thèse, Université de Strasbourg.

Arsac J. (1977). *La construction de programmes structurés*. Dunod, Paris.

Arsac J. (1980). *Premières leçons de programmation*, Cedic- Nathan, Paris.

Arsac J. (1987). *Des pédagogies pour l'option informatique*, Pratiques et savoir-faire des élèves de l'option informatique, 4, p.1-13.

Arsac J. (1989). *La didactique de l'informatique : un problème ouvert ?* Contribution au Colloque francophone sur la didactique de l'informatique (1988). Actes publiés par l'E.P.I.

Artigue M. (1991). *Epistémologie et didactique*, Recherches en Didactique des Mathématiques. 10 (2/3), p. 241-285

Auriol J.C. (2007). *Formation à la statistique par la pratique d'enquêtes par questionnaires et la simulation : étude didactique d'une expérience d'enseignement dans un département d'IUT* –Thèse. Lyon, Université Lumière Lyon 2.

Bachelard G. (1938). *La formation de l'esprit scientifique*. Librairie philosophique Vrin, Paris.

Balacheff N. (1995 a). *Conception, Connaissance et Concept*, Didactique et technologies cognitives en mathématiques. Séminaires 1994-1995. (p. 219-244). Dida Tech. Laboratoire de structures discrètes et de didactique. CNRS-IMAG: Université Joseph Fourier, Grenoble.

Balacheff N. (1995 b). *Conception, propriété du système sujet/milieu*, Noirfalise R., Perrin-Glorian M.-J. (ed) Actes de la VIIe Ecole d'été de didactique des mathématiques (p. 215-229). Clermont-Ferrand : IREM de Clermont-Ferrand.

Baron G.L., **Bruillard** E. (2001). *Une didactique de l'informatique ?*, la Revue Française de Pédagogie, n° 135, p. 163-172

Baron G.L. (1987). *La constitution de l'informatique comme discipline scolaire, le cas des lycées*, Thèse en sociologie de l'Education, Université René Descartes, Paris

Bordier J. (1991). *Un modèle didactique, utilisant la simulation sur ordinateur, pour l'enseignement de la probabilité* – Thèse. Paris, Université Paris VII.

Boulanger J. (2009). *Décisions, préférences et probabilités : deux expérimentations confrontant des élèves à des situations mettant en scène le hasard*. Thèse. Paris, Université Paris Descartes.

Brousseau G. (1971). *Vers un enseignement des probabilités à l'école élémentaire*. Enseignement élémentaire des mathématiques, publié par l'IREM de Bordeaux.

Brousseau G. (1998). *Théorie des situations didactiques*. Grenoble, La Pensée Sauvage éd.

Brousseau G. (2003). *Situations fondamentales et processus génétiques*. Actes de la XIIe école d'été de didactique des mathématiques (août 2003), ARDM, Grenoble, Éditions La Pensée sauvage.

Chevallard Y., **Wozniak** F. (2003 a). *Enseigner la statistique au secondaire. Entre genre prochain et différence spécifique*. Cours donné à la XIIe école d'été de didactique des mathématiques (Corps, 20-29 août 2003). Paru dans Mercier, A. & Margolinas, C. (Eds), Balises pour la didactique des mathématiques, La Pensée sauvage, Grenoble, p. 195-218.

Chevallard Y., **Wozniak** F. (2003 b). *Enseigner la statistique : des mathématiques mixtes pour penser la variabilité*. Compte rendu d'un atelier animé dans le cadre de la XIIe école d'été de didactique des mathématiques (Corps, 20-29 août 2003)

Chevallard Y. (1978). *Notes pour la didactique de la statistique*, IREM, Aix-Marseille.

Chevallard Y. (1986). *La transposition didactique : du savoir savant au savoir enseigné*. La pensée sauvage éditions.

Chevallard Y. (1992) *Intégration et viabilité des objets informatiques dans l'enseignement des mathématiques*, Contribution à l'ouvrage dirigé par B. Cornu, L'ordinateur pour enseigner les mathématiques, PUF, 1992, p. 183-203.

Chevallard Y. (1992). *Concepts fondamentaux de la didactique : perspectives apportées par une approche anthropologique*, Recherches en Didactique des Mathématiques 12 (1), p. 73-112

Chevallard Y. (1994). *Les processus de transposition didactique et leur théorisation*, Contribution à l'ouvrage dirigé par G. Arsac, Y. Chevallard, J.-L. Martinand, Andrée Tiberghien (éds), La transposition didactique à l'épreuve, La Pensée sauvage, Grenoble, p. 135-180.

Chevallard Y. (1998). *A propos des TICE : transmission et appropriation du savoir, nouveaux rôles de l'enseignant, organisation de l'établissement*, Conférence donnée à l'université d'été Les TICE : vers une transformation des pratiques pédagogiques et de l'organisation de l'établissement (Toulouse, 26-28 août 1998).

Chevallard Y. (1998). *Analyse des pratiques enseignantes et didactique des mathématiques : l'approche anthropologique*. Cours donné à l'université d'été Analyse des pratiques enseignantes et didactique des mathématiques, La Rochelle, 4-11 juillet 1998 ; paru dans les actes de cette université d'été, IREM de Clermont-Ferrand, p. 91-120.

Chevallard Y. (2003). *Approche anthropologique au savoir et didactique des mathématiques*, contribution à l'ouvrage « Rapport au savoir et didactiques », Education et Sciences, éditions Fabert.

Chevallard Y. (2007). *Une épreuve expérimentale de mathématiques ? Réponse à une question de Michèle Artigue dans le cadre d'un forum du site EducMath*. Mise en ligne : 21 septembre 2007 : http://educmath.inrp.fr/Educmath/en-debat/epreuve-pratique

Cohen J. (1963). *Hasard, adresse et chance : la psychologie du pari et du jeu*. Presses Universitaires de France.

Cohen J., **Hansel** M. (1957) *Risque et jeux : probabilités subjectives*. Delachaux et Niestlé.

Cori R., **Lévy** J.J. (2011). *Algorthmes et programmation*. Cours de tronc commun. Ecole Polytechnique.

http://www.enseignement.polytechnique.fr/profs/informatique/Jean-Jacques.Levy/poly/

Cournot A. (1843). *Exposition de la théorie des chances et des probabilités*. Paris, Hachette.

Courtebras B. (2008 a) *Formes de rapports au calcul des probabilités*. Collection EPU. Sciences humaines et sociales. Publibook, Paris.

Courtebras B. (2006). *A l'école des probabilités*. Didactiques Mathématiques, Presses Universitaires de Franche-Comté.

Courtebras B. (2008). *Mathématiser le hasard. Une histoire du calcul des probabilités.* Collection « Culture scientifique ». Vuibert, Paris.

Coutanson B. (2010). *La question de l'éducation statistique et de la formation de l'esprit statistique à l'école primaire en France*. Thèse, Université Lumière Lyon 2.

Coutinho C. (2001). *Introduction aux situations aléatoires dès le collage: de la modélisation à la simulation d'expériences de Bernoulli dans l'enseignement informatique Cabri Géomètre II* – Thèse. Grenoble, Université Joseph Fourrier.

Dacunha-Castelle D. (1996). *Chemins de l'aléatoire : le hasard et le risque dans la société moderne*. Flammarion.

Dacunha-Castelle D. (2009). *Le difficile chemin de l'aléatoire dans les programmes*, bulletin n°484 de l'A.P.M.E.P., octobre 2009.

Dagnelie P. (2010). *Dans l'enseignement de la statistique, la simulation a des limites*. « Libres propos » dans la revue numérique « Statistique et Enseignement » de la Société Française de Statistique (SfdS), 1(2), 67-68, http://www.statistique-et-enseignement.fr

Darmois G. (1956) *La formation probabiliste et statistique de l'ingénieur*, Revue de statistique appliquée, tome 4, N°2 (1956), p. 13-16.

Delahaye J.P. (2005). *Le trésor et les Sophies*, Pour La Science, n° 336.

Dhieb M. (2009). *Contribution à l'introduction des probabilités au collège : rapports d'élèves à quelques notions probabilistes* – Thèse. Paris, Université Paris Descartes.

Dijkstra E.W., **Dahl** O.J., **Moare** C.A.R. (1972). *Structured programming*, Academic Press, London.

Dijkstra E.W. (1968). *GOTO statements considered as harmful*. Comm. ACM 11, 3, mars 1968, p. 147-148.

Dijkstra E.W.(1976). *A discipline of programming*. Prentice Hall

Dodge Y. (1993). Statistique. Dictionnaire encyclopédique. Dunod, Paris.

Fine J. (2010). *Probabilités et statistique inférentielle. Approche sondage versus approche modèle*, « Recherches et Perspectives » dans la revue numérique « Statistique et Enseignement » de la Société Française de Statistique (SfdS), 1(2), p. 67-68, http://www.statistique-et-enseignement.fr

Fischbein E., **Pampou** I., **Minzat** I. (1967). *L'intuition probabiliste chez l'enfant*. Enfance, tome 20 n° 2, p. 193-208.

Fischbein E., **Pampou** I., **Minzat** I. (1969). *Initiation aux probabilités à l'école élémentaire*, Educational Studies in Mathematics , Volume 2. p. 16-31.

Fischbein E. (1975). *The intuitive source of probabilistic thinking in children*. Reidel.

Flajolet, P., **Parizot** E. (2004). *Qu'est-ce qu'un algorithme ?* site « Interstices », février 2004 : http://interstices.info/jcms/c_5776/quest-ce-quun-algorithme.

Girard J.C., **Henry** M. (2005). *Modélisation et simulation en classe, quel statut didactique ?* dans Statistique au lycée, volume 1, brochure A.P.M.E.P. n° 156

Glayman M., **Varga** T. (1973). *Les probablités à l'école,* Cedic, Paris.

Gries D. (1981). *The science of programming*. Springer Verlag. Berlin.

Henry M. (1999). *L'introduction des probabilités au lycée : un processus de modélisation comparable à celui de la géométrie*. REPERES-IREM, n° 36.

Henry M. (2001). *Notion d'expérience aléatoire, vocabulaire et modèle probabiliste*, Autour de la modélisation en probabilités, p.161-172, Commission Inter-IREM Statistique et Probabilités , collection Didactiques, Presses Universitaires de Franche-Comté, Besançon.

Kahane J.P. (2000). Commission de réflexion sur l'enseignement des mathématiques. Rapport d'étape sur l'informatique et l'enseignement des mathématiques. Décembre 2000.

Kahneman D., **Twersky** A. (1972). *Subjective probability : a judgement of representativeness*. Cognitive Psychologie, 3.

Kahneman D., **Twersky** A. (1982). *On the study of statistical intuitions*. Cognition, 11.

Knuth D. (1968). *The art of computer programming*, vol 1, Addison Wesley, 1968.

Lahanier-Reuter D. (1999) . *Conceptions du hasard et enseignement des probabilités et statistiques*. Presses Universitaires de France.

Lecoutre M.P. (1985). *Jugements probabilistes chez les adultes : pratique des jeux de hasard et formation en théorie des probabilités.* Bulletin de psychologie, Tome 38, n° 372.

Martin T. (2003). *Probabilité et certitude*, p.119-134 dans « Probabilités subjectives et rationalité de l'action » (travail collectif sous la direction de Martin), CNRS Editions, Paris.

Martin T. (2007). *L'épistémologie probabiliste de Cournot,* dans « La société du probable, les mathématiques sociales après Augustin Cournot », p. 37- 62. Albin Michel

Maury S. (1985). *Influence de la question dans une épreuve relative à la notion d'indépendance*, Educationale Studies in Mathematics, 16, p 283-301.

Maury S. (1986). *Contribution à l'étude didactique de quelques notions de probabilité et de combinatoire à travers la résolution de problèmes* – Thèse. Université des Sciences et Techniques du Languedoc, Montpellier.

Maury S. (1987). *Procédures dans la résolution de problèmes probabilistes*, dans « Didactiques et acquisition des connaissances scientifiques », pp. 293-304, La Pensée Sauvage, Grenoble.

Maury S., **Caillot**, M. (2003). *Quand les didactiques rencontrent le rapport au savoir*, dans « Rapport au savoir et didactiques », Education et Sciences, éditions Fabert.

Morand B. (2009). *Quels savoir-faire faut-il cultiver pour les candidats aux métiers de l'informatique ?* dans : « Informatique et progiciels en éducation et en formation : continuités et perspectives », ouvrage sous la direction de Georges-Louis BARON, Éric BRUILLARD et Luc- Olivier POCHON. I.N.R.P., Lyon 2009.

Nabbout M. (2006). *Enseignement des probabilités en classe terminale au Liban : étude de représentations et de pratiques de professeurs dans des situations aménagées* – Thèse de doctorat. Paris, Université René Descartes.

Naur P. (1969). *Programming by action clusters*. BIT 9, 3, p. 250-258.

Neveu J. (1970). *Bases mathématiques du calcul des probabilités*, Masson et Cie, Paris.

Nguyen C.T. (2005). *Etude didactique de l'introduction d'éléments d'algorithmique et de programmation dans l'enseignement mathématique secondaire à l'aide d'une calculatrice.* Thèse. Université Joseph Fourier, Grenoble.

Nivat M. (2009). *Vaches et informatique* dans : « Informatique et progiciels en éducation et en formation : continuités et perspectives », ouvrage sous la direction de Georges-Louis BARON, Éric BRUILLARD et Luc-Olivier POCHON. I.N.R.P., Lyon 2009.

Oriol J.C. (2007). *Formation à la statistique par la pratique d'enquêtes par questionnaires et la simulation : étude didactique d'une expérience d'enseignement dans un département d'IUT*, Thèse en Sciences de l'éducation, Université Lumière, Lyon 2.

Pair C. (1988). *Je ne sais (toujours) pas enseigner la programmation*, dans la revue « Informatiques » n° 2, pp 5-14

Pair C. (1989). *L'apprentissage de la programmation*, Contribution au Colloque francophone sur la didatique de l'informatique (1988). Actes publiés par l'E.P.I.

Parzysz B. (2003). *L'enseignement de la statistique et des probabilités en France : évolution au cours d'une carrière d'enseignant (période 1965-2002)*, Probabilités au lycée, brochure 143, p 9-34, A.P.M.E.P.

Parzysz B. (2007). *Expérience aléatoire et simulation : le jeu de croix ou pile. Relecture actuelle d'une expérimentation déjà un peu ancienne.* REPERES-IREM n°66

Parzysz B. (2009). *De l'expérience aléatoire au modèle, via la simulation*. REPERES-IREM n°74, janvier 2009.

Pascal B. (1670). *Œuvres complètes*, l'Intégrale / Seuil.

Piaget J., **Inhelder** B. (1951). *La génèse de l'idée de hasard chez l'enfant*. PUF, Paris.

Régnier J.C. (2002). *A propos de la formation en statistique. Approches praxéologiques et épistémologiques de questions du champ de la didactique de la statistique*, Revue du Centre de recherche en Education , 22/23, 2003, Université Jean-Monnet, Saint Etienne.

Regnier J.C. (2005). *Formation de l'esprit statistique et raisonnement statistique. Que peut-on attendre de la didactique de la statistique ?* , Actes du séminaire national de Didactique des Mathématiques. Année 2005. Editeurs : ARDM & IREM de Paris 7 (p.13-37).

Régnier JC., **Oriol** J.C. (2009). *Quelques obstacles rencontrés dans l'apprentissage de l'analyse statistique implicative*. 41 èmes Journées de Statistique, SfdS, Bordeaux 2009

Renyi A. (1966). *Calcul des probabilités*, suivi de *Introduction à la théorie de l'information*, (réédition 1992), Jacques Gabay, Paris.

Robert C. (2003). *Contes et décomptes de la statistique : une introduction par l'exemple*, Edition Vuibert.

Robert A., **Rogalski** J. , **Samurcay** R. (1987). *Enseigner des méthodes*, Cahiers de Didactique des Mathématiques, 38, IREM, Université Paris VII.

Roditi E. (2009). *Un tableur graphique pour enseigner les statistiques en sciences humaines et sociales*, dans : « Informatique et progiciels en éducation et en formation : continuités et perspectives » ; sous la direction de Georges-Louis BARON, Éric BRUILLARD et Luc-Olivier POCHON. I.N.R.P., Lyon 2009, p. 257-275.

Rogalski J., **Samurcay** R., **Hoc** J.M . (1988). *L'apprentisage des méthodes de programmation comme méthode de résolution de problème*, « Le Travail Humain », 51.

Rogalski J. (1986). *Pour une pédagogie de l'informatique*, Bulletin de l'EPI n° 42, juin 1986, p 105-109

Rogalski J. (1987). *Acquisition de savoirs et savoir-faire en informatique*, Cahiers de Didactique des Mathématiques, 43, IREM, Université Paris VII,

Rogalski J. (1989). *Enseignement de méthodes de programmation dans l'initiation à l'informatique*. Contribution au Colloque francophone sur la didatique de l'informatique (1988). Actes publiés par l'E.P.I.

Rogalski J.,**Vergnaud** G. (1987). *Didactique de l'informatique et acquisitions cognitives en programmation*, Psychologie Française, 4, p. 267-273.

Schwartz C. (1998). *L'empereur et la girafe, initiation à la statistique*. Diderot éditeur, Arts et Sciences.

Schwartz C. (2011) *A propos du document ressources pour la classe de première statistiques et probabilités*, article sur le site « statistix », juin 2011
http://www.statistix.fr/spip.php?article89

Shafer G. (2007). *Du principe de Cournot au marché efficient*. Dans « dans « La société du probable, les mathématiques sociales après Augustin Cournot », p. 83 à 132. Albin Michel

Vergnaud G. (1991). *La théorie des champs conceptuels*, R.D.M. vol. 10/2.3.

Walliser B. [1997]. *Systèmes et modèles. Introduction critique à l'analyse de systèmes*. Editions du Seuil, Paris.

Wozniak F. (2005). *Conditions et contraintes de l'enseignement de la statistique en classe de seconde générale. Un repérage didactique*. Thèse de doctorat, Université Claude-Bernard Lyon 1.

Zaki M. (1990). *Traitement de problèmes de probabilités en situation de simulation*, Thèse, Université de Strasbourg.

Documents numériques cités dans la thèse :

[1] « Ressources pour la classe de seconde, algorithmique », (« programmes 2010 »)
http://media.eduscol.education.fr/file/Programmes/17/8/Doc_ress_algo_v25_109178.pdf

[2] « Ressources pour la classe de première générale et technologique : Statistiques et probabilités », (« programmes 2010 »)
http://media.eduscol.education.fr/file/Mathematiques/12/7/LyceeGT_Ressources_Maths_1_Statistiques_Probabilites_182127.pd

[3] « Ressources pour la classe de seconde, Probabilités et Statistiques», (« programmes 2010 »)
http://media.eduscol.education.fr/file/Programmes/17/9/Doc_ressource_proba-stats_109179.pdf

[4] « histoire » de l'option informatique des lycée :
http://edutice.archives-ouvertes.fr/docs/00/56/45/59/HTML/h10oi_jb1.htm

Documents d'accompagnement du programme de seconde , « programmes 2000 »
http://www2.cndp.fr/gtd_maths/pdf/ESEMA003.pdf

Sites évoqués dans la thèse

Interstices :
Ce site *été créé en 2004 à l'initiative de l'INRIA, Institut national de recherche en informatique et en automatique. Il se développe dans un partenariat entre l'INRIA, le CNRS, les Universités et l'ASTI, Association Française des Sciences et Technologies de l'Information.*
http://interstices.info/jcms/jalios_5127/accueil

Site de l'A.P.M.E.P. : Association des Professeurs de Mathématiques de l'Enseignement Public.
http://www.apmep.asso.fr/

site de l'E.P.I. : « association Enseignement Public & Informatique »
http://www.epi.asso.fr

Statistix : ce site est présenté sur la page d'accueil par : *« Centre de ressources, lieu de partage et de mutualisation pour l'enseignement de la statistique. Pour qui ? Les enseignants des écoles, des collèges et des lycées de toutes les disciplines. »*
http://statistix.fr/

Site de la Société Française de Statistisque (SfdS) :

http://www.sfds.asso.fr/

Site de la Société de Mathématique de France (SMF) :

http://smf.emath.fr/Enseignement/CommissionKahane

Eduscol : site du ministère de l'Education Nationale :

http://eduscol.education.fr/

ANNEXES

Les documents annexés se rapportant à un même chapitre sont regroupés.

Ces documents sont ici référencés par un numéro et un titre.

Nous présentons ci dessous un tableau expliquant l'agencement de ces documents annexés :

numéro	titre	chapitre
1	programme de seconde (« programmes 2000 ») : extrait concernant la statistique et les probabilités	3
2	programme de première S (« programmes 2000 ») : extrait concernant la statistique et les probabilités	3
3	programme de terminale S (« programmes 2000 ») : extrait concernant la statistique et les probabilités	3
4	sujets du baccalauréat sur l' « adéquation » (test d'hypothèse)	3
5	3 sujets de l'épreuve pratique expérimentale	3
6	programme de seconde (« programmes 2010 ») : extrait concernant la statistique et les probabilités	3
7	« programmes 2010 » ressources pour la classe de seconde : algorithmique (extrait)	11
8	« programmes 2010 » ressources pour la classe de première : statistiques et probabilités (extrait)	11
9	programme de première S (« programmes 2010 ») : extrait concernant la statistique et les probabilités	11
10	Note de l'Inspection Générale de mathématiques aux concepteurs de sujets	13
11	Valeurs du χ^2 pour les écarts observés dans divers tableaux	17
12	Transcription des réponses des élèves à la question B. (« Comment aider Monsieur Dupond ? »)	18
13	Projets de programme de statistique et probabilités pour les terminales S (« programmes 2010) extrait concernant la statistique et les probabilités	21

Annexes du chapitre 3

Annexe 1 : programme de seconde (« programmes 2000 »)

STATISTIQUE

Rappel des programmes antérieurs :

SIXIÈME	CINQUIÈME	QUATRIÈME	TROISIÈME
Exemples conduisant à lire et établir des relevés statistiques sous forme de tableaux ou de représentations graphiques, éventuellement en utilisant un ordinateur.	*Lecture, interprétation, représentations graphiques de séries statistiques. Diagrammes à barres, diagrammes circulaires. Classes, effectifs. Fréquences.*	*Effectifs cumulés, fréquences cumulées. Moyennes pondérées. Initiation à l'usage des tableurs-grapheurs. Valeur approchée de la moyenne d'une série statistique regroupée en classes d'intervalles.*	*Caractéristiques de position d'une série statistique. Approche de caractéristiques de dispersion d'une série statistique. Initiation à l'utilisation des tableurs-grapheurs en statistique.*

En seconde le travail sera centré sur :
- la réflexion conduisant au choix de résumés numériques d'une série statistique quantitative ;
- la notion de fluctuation d'échantillonnage vue ici sous l'aspect élémentaire de la variabilité de la distribution des fréquences ;
- la simulation à l'aide du générateur aléatoire d'une calculatrice. La simulation remplaçant l'expérimentation permet, avec une grande économie de moyens, d'observer des résultats associés à la réalisation d'un très grand nombre d'expériences. On verra ici la diversité des situations simulables à partir d'une liste de chiffres.

L'enseignant traitera des données en nombre suffisant pour que cela justifie une étude statistique ; il proposera des sujets d'étude et des simulations en fonction de l'intérêt des élèves, de l'actualité et de ses goûts.
Les notions de fluctuation d'échantillonnage et de simulation ne doivent pas faire l'objet d'un cours. L'élève pourra se faire un "cahier de statistique" où il consignera une grande partie des traitements de données et des expériences de simulation qu'il fait, des raisons qui conduisent à faire des simulations ou traiter des données, l'observation et la synthèse de ses propres expériences et de celles de sa classe. Ce cahier sera complété en première et terminale et pourra faire partie des procédures d'évaluation annuelle.
En classe de première et de terminale, dans toutes les filières, on réfléchira sur la synthèse des données à l'aide du couple (moyenne, écart-type) qui sera vu à propos de phénomènes aléatoires gaussiens et par moyenne ou médiane et intervalle inter-quartile sinon. On amorcera une réflexion sur le problème de recueil des données et sur la notion de preuve statistique ; on fera un lien entre statistique et probabilité.
L'enseignement de la statistique sera présent dans toutes les filières mais sous des formes diverses.

CONTENUS	CAPACITÉS ATTENDUES	COMMENTAIRES
Résumé numérique par une ou plusieurs mesures de tendance centrale (moyenne, médiane, classe modale, moyenne élaguée) et une mesure de dispersion (on se restreindra en classe de seconde à l'étendue).	Utiliser les propriétés de linéarité de la moyenne d'une série statistique. Calculer la moyenne d'une série à partir des moyennes de sous-groupes. Calcul de la moyenne à partir de la distribution des fréquences.	L'objectif est de faire réfléchir les élèves sur la nature des données traitées, et de s'appuyer sur des représentations graphiques pour justifier un choix de résumé. On peut commencer à utiliser le symbole Σ. On commentera quelques cas où la médiane et la moyenne diffèrent sensiblement. On remarquera que la médiane d'une série ne peut se déduire de la médiane de sous séries. Le calcul de la médiane nécessite de trier les données, ce qui pose des problèmes de nature algorithmique.

Définition de la distribution des fréquences d'une série prenant un petit nombre de valeurs et de la fréquence d'un événement. Simulation et fluctuation d'échantillonnage.	Concevoir et mettre en œuvre des simulations simples à partir d'échantillons de chiffres au hasard.	La touche "random" d'une calculatrice pourra être présentée comme une procédure qui, chaque fois qu'on l'actionne, fournit une liste de n chiffres (composant la partie décimale du nombre affiché). Si on appelle la procédure un très grand nombre de fois, la suite produite sera sans ordre ni périodicité et les fréquences des dix chiffres seront sensiblement égales. Chaque élève produira des simulations de taille n (n allant de 10 à 100 suivant les cas) à partir de sa calculatrice ; ces simulations pourront être regroupées en une simulation ou plusieurs simulations de taille N, après avoir constaté la variabilité des résultats de chacune d'elles. L'enseignant pourra alors éventuellement donner les résultats de simulation de même taille N préparées à l'avance et obtenues à partir de simulations sur ordinateurs.

THÈMES D'ÉTUDE

Pour chacun des chapitres, le professeur choisira, pour l'ensemble des élèves ou pour certains seulement en fonction de leurs centres d'intérêt, un ou plusieurs thèmes d'étude dans la liste ci-dessous.

Statistique

- Simulations d'un sondage ; à l'issue de nombreuses simulations, pour des échantillons de taille variable, on pourra introduire la notion de fourchette de sondage, sans justification théorique. La notion de niveau de confiance 0,95 de la fourchette peut être introduite en terme de "chances" (il y a 95 chances sur 100 pour que la fourchette contienne la proportion que l'on cherche à estimer) ; on pourra utiliser les formules des fourchettes aux niveaux 0,95, 0,90 et 0,99 pour une proportion observée voisine de 0,5 afin de voir qu'on perd en précision ce qu'on gagne en niveau de confiance. On incitera les élèves à connaître l'approximation usuelle de la fourchette au niveau de confiance 0,95, issue d'un sondage sur n individus (n>30) dans le cas où la proportion observée $\hat{p}$ est comprise entre 0,3 et 0,7, à savoir : $[\hat{p} - 1/\sqrt{n} \,;\, \hat{p} + 1/\sqrt{n}]$.

- Simulations de jeux de pile ou face : distribution de fréquences du nombre maximum de coups consécutifs égaux dans une simulation de 100 ou 200 lancers d'une pièce équilibrée; distribution de fréquences du gain sur un jeu d'au plus dix parties où on joue en doublant la mise (ou en la triplant) tant qu'on n'a pas gagné. On pourra aussi faire directement l'expérience avec des pièces pour bien faire sentir la notion de simulation...
- Simulations du lancer de deux dés identiques et distribution de la somme des faces. On pourra aussi faire directement l'expérience avec des dés pour bien faire sentir la notion de simulation...
- Simulations de promenades aléatoires sur des solides ou des lignes polygonales, fluctuation du temps et estimation du temps moyen mis pour traverser un cube ou pour aller d'un sommet donné à un autre sommet donné d'une ligne polygonale.
- Simulations de naissances : distribution du nombre d'enfants par famille d'au plus quatre enfants lorsqu'on s'arrête au premier garçon, en admettant que pour chaque naissance, il y a autant de chances que ce soit un garçon ou une fille.

****

Annexe 2 : programme de première S (« programmes 2000 »)

PROBABILITÉS ET STATISTIQUE

La partie du programme consacrée aux probabilités et à la statistique est centrée:
- sur la mise en place d'éléments de base indispensables pour comprendre ou pratiquer la statistique partout où elle est présente,
- sur l'acquisition de concepts de probabilité permettant de comprendre et d'expliquer certains faits simples observés expérimentalement ou par simulation.

Le programme de la classe de première introduit quelques outils descriptifs nouveaux:
- les diagrammes en boîtes qui permettent d'appréhender aisément certaines caractéristiques des répartitions des caractères étudiés et qui complètent la panoplie des outils graphiques les plus classiquement utilisés;
- deux mesures de dispersion: l'écart-type et l'intervalle interquartile.

Ces éléments de statistique pourront notamment être travaillés pour des séries construites à partir de séries simulées; on rencontre ainsi des répartitions variées et on prépare la notion d'estimateur. Cette partie descriptive ne doit pas faire l'objet de longs développements numériques, ni être déconnectée du reste du programme de probabilité et statistique.

CONTENUS	MODALITÉS DE MISE EN ŒUVRE	COMMENTAIRES
Statistique Variance et écart-type. Diagramme en boîte; intervalle interquartile. Influence sur l'écart-type et l'intervalle interquartile d'une transformation affine des données.	On cherchera des résumés pertinents et on commentera les diagrammes en boîtes de quantités numériques associées à des séries simulées ou non. On observera l'influence des valeurs extrêmes d'une série sur l'écart-type ainsi que la fluctuation de l'écart-type entre séries de même taille. L'usage d'un tableur ou d'une calculatrice permettent d'observer dynamiquement et en temps réel, les effets des modifications des données.	L'objectif est de résumer une série par un couple (mesure de tendance centrale; mesure de dispersion). Deux choix usuels sont couramment proposés: le couple (médiane; intervalle interquartile), robuste par rapport aux valeurs extrêmes de la série, et le couple (moyenne ; écart-type). On démontrera que la moyenne est le réel qui minimise $\sum(x_i-x)^2$, alors qu'elle ne minimise pas $\sum \mid x_i-x \mid$. On notera s l'écart-type d'une série, plutôt que σ, réservé à l'écart-type d'une loi de probabilité.
Probabilités Définition d'une loi de probabilité sur un ensemble fini. Espérance, variance, écart-type d'une loi de probabilité. Probabilité d'un événement, de la réunion et de l'intersection d'événements. Cas de l'équiprobabilité. Variable aléatoire, loi d'une variable aléatoire, espérance, variance, écart-type. Modélisation d'expériences aléatoires de référence (lancers d'un ou plusieurs dés ou pièces discernables ou non, tirage au hasard dans une urne, choix de chiffres au hasard, etc.).	Le lien entre loi de probabilité et distributions de fréquences sera éclairé par un énoncé vulgarisé de la loi des grands nombres. On expliquera ainsi la convergence des moyennes vers l'espérance et des variances empiriques vers les variances théoriques; on illustrera ceci par des simulations dans des cas simples. On pourra aussi illustrer cette loi avec les diagrammes en boîtes obtenus en simulant par exemple 100 sondages de taille n, pour $n = 10; 100; 1000$. On simulera des lois de probabilités simples obtenues comme images d'une loi équirépartie par une variable aléatoire (sondage, somme des faces de deux dés, etc.).	On pourra par exemple choisir comme énoncé vulgarisé de la loi des grands nombres la proposition suivante: *Pour une expérience donnée, dans le modèle défini par une loi de probabilité P, les distributions des fréquences calculées sur des séries de taille n se rapprochent de P quand n devient grand.* On indiquera que simuler une expérience consiste à simuler un modèle de cette expérience. La modélisation avec des lois ne découlant pas d'une loi équirépartie est hors programme. On évitera le calcul systématique et sans but précis de l'espérance et de la variance de lois de probabilité.

****

Annexe 3 : programmes de terminale S (« programmes 2000 »)

(document CNDP, réédition novembre 2005)

II. 3 Probabilités et statistique

Aprés avoir introduit en classe de seconde la nature du questionnement statistique à partir de travaux sur la fluctuation d'échantillonnage, on poursuit ici la présentation entreprise en première des concepts fondamentaux de probabilité dans le cas fini avec la notion de conditionnement et d'indépendance et l'étude de quelques lois de probabilité.

On vise aussi, en complément à l'usage des simulations introduit dès la seconde, une première sensibilisation à d'autres classes de problèmes, notamment celui de l'adéquation d'une loi de probabilité à des données expérimentales.

CONTENUS	MODALITÉS DE MISE EN ŒUVRE	COMMENTAIRES
Conditionnement et indépendance		
Conditionnement par un événement de probabilité non nulle puis indépendance de deux événements. Indépendance de deux variables aléatoires.	On justifiera la définition de la probabilité de B sachant A, notée $P_A(B)$, par des calculs fréquentiels. On utilisera à bon escient les représentations telles que tableaux, arbres, diagrammes, etc., efficaces pour résoudre des problèmes de probabilités.	Un arbre de probabilité correctement construit constitue une preuve.
Formule des probabilités totales.	Application à la problématique des tests de dépistage en médecine et à la loi de l'équilibre génétique lors d'appariements au hasard.	Les élèves doivent savoir appliquer sans aide la formule des probabilités totales dans des cas simples.
Statistique et modélisation. Expériences indépendantes. Cas de la répétition d'expériences identiques et indépendantes.	Application aux expériences de références vues en seconde et première (dés, pièces, urnes, etc.).	On conviendra, en conformité avec l'intuition, que pour des expériences indépendantes, la probabilité de la liste des résultats est le produit des probabilités de chaque résultat.
Lois de probabilité		
Exemples de lois discrètes Introduction des combinaisons, notées $\binom{n}{p}$ Formule du binôme.	On introduira la notation n !. L'élève devra savoir retrouver les formules : $\binom{n}{p} = \binom{n-1}{p-1} + \binom{n-1}{p}$ $\binom{n}{p} = \binom{n}{n-p}$	Le symbole $\binom{n}{p}$ peut être désigné par la locution « p parmi n ». Pour les dénombrements intervenant dans les problèmes, on en restera à des situations élémentaires résolubles à l'aide d'arbres, de diagrammes ou de combinaisons.
Loi de Bernoulli, loi binomiale ; espérance et variance de ces lois.	On appliquera ces résultats à des situations variées.	La formule donnant l'espérance sera conjecturée puis admise ; la formule de la variance sera admise.
Exemples de lois continues Lois continues à densité : – loi uniforme sur [0,1] ; – loi de durée de vie sans vieillissement.	Application à la désintégration radioactive : loi exponentielle de désintégration des noyaux.	Ce paragraphe est une application de ce qui aura été fait en début d'année sur l'exponentielle et le calcul intégral.
Statistique et simulation.	Étude d'un exemple traitant de l'adéquation de données expérimentales à une loi équirépartie.	L'élève devra être capable de poser le problème de l'adéquation à une loi équirépartie et de se reporter à des résultats de simulation qu'on lui fournit. Le vocabulaire des tests (test d'hypothèse, hypothèse nulle, risque de première espèce) est hors programme.

Annexe 4 : 3 exercices de baccalauréat portant sur l' « adéquation » de la loi équirépartie

Antilles-Guyane, septembre 2004

EXERCICE 4 **5 points**

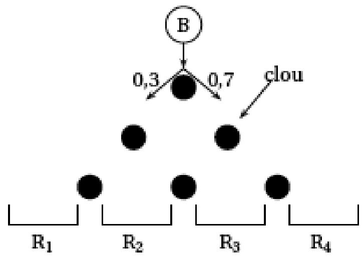
Pour chacune des trois questions, la totalité des points sera donnée si la réponse est correctement justifiée.
Les trois questions sont indépendantes.

1. La probabilité pour un individu d'une population d'être atteint d'une maladie M est égale à 0,003. Un test de dépistage, pour cette maladie, a été réalisé ; avec ce test, on peut dire que
   - si une personne est atteinte de la maladie M, le test est positif dans 50 % des cas ;
   - le test est positif pour 3 % des personnes saines.

   Quelle est à 0,01 près la probabilité d'avoir la maladie M lorsque le test est positif ?

   ☐ 0,95 ☐ 0,9 ☐ 0,15 ☐ 0,05

2. On considère une planche à clous de ce type :



   On lance une boule B du haut de la planche, elle tombe alors dans l'un des quatre récipients notés R_1, R_2, R_3 et R_4. À chaque étape, la bille a une probabilité de 0,3 d'aller vers la gauche et 0,7 d'aller vers la droite (gauche et droite relatives à l'observateur).
   On note p_1 la probabilité que la bille tombe dans le bac R_1 ou dans le bac R_3 et p_2 la probabilité que la bille tombe dans le bac R_2 ou dans le bac R_4.
   Que valent p_1 et p_2 ?

   ☐ $p_1 = p_2 = 0,5$ ☐ $p_1 = 0,216$ et $p_2 = 0,784$
   ☐ $p_1 = 0,468$ et $p_2 = 0,532$ ☐ $p_1 = 0,468$ et $p_2 = 0,432$.

3. Les 1 000 premières décimales de π sont données ici par un ordinateur :

   1415926535 8979323846 2643383279 5028841971 6939937510
   5820974944 5923078164 0628620899 8628034825 3421170679
   8214808651 3233066470 9384460959 0582235725 3594085234
   8111745028 4102701930 5211055596 4462294895 4930301964
   4288109756 6593344612 8475648233 7867831652 7120190914
   5648566923 4603486534 5432664825 3393607260 2491412737
   2450700660 6315580574 8815209209 6282925409 1715364367
   8925903600 1133053054 8820466525 3841469519 4151160943
   3057270365 7595919530 9218611738 1932611793 1051185480
   7446297996 2749567355 8857527240 9122793318 3011949129
   8336733624 4065664308 6025394946 3952247371 9070217986
   0943702770 5392171762 9317675238 4674818467 6691051320
   0056812714 5263560827 7857753427 9778900917 3637178721
   4684409012 2495343054 6549585371 0507922796 8925892354
   2019956112 1290219608 6403441815 9813629774 7713099605
   1870721134 9999998372 9780499510 5973173281 6096318599
   0244594553 4690830264 2522300253 3446850352 6193110017
   1010003137 8387528865 8753320830 1420617177 6691473035
   9825349042 8755460731 1595620633 8235378759 3751957781
   8577805321 7122600661 3001927876 6111959092 1642019894

   En groupant par valeurs entre 0 et 9 ces décimales, on obtient le tableau suivant :

Valeurs	0	1	2	3	4	5	6	7	8	9
Occurrences	93	116	102	102	94	97	94	95	101	106

Avec un tableur, on a simulé 1 000 expériences de 1 000 tirages aléatoires d'un chiffre compris entre 0 et 9.

Pour chaque expérience, on a calculé $d^2 = \sum_{k=0}^{k=9} (f_k - 0,1)^2$ où f_k représente, pour l'expérience, la fréquence observée du chiffre k.
On a alors obtenu une série statistique pour laquelle on a calculé le premier et neuvième décile (d_1 et d_9), le premier et troisième quartile (Q_1 et Q_3) et la médiane (Me) :
$d_1 = 0,000\ 422$; $Q_1 = 0,000\ 582$; Me $= 0,000\ 822$; $Q_3 = 0,001\ 136$; $d_9 = 0,001\ 45$.

En effectuant le calcul de d_2 sur la série des 1 000 premières décimales de π, on obtient :

☐ 0,000 456 ☐ 0,004 56 ☐ 0,000 314

Un statisticien découvrant le tableau et ignorant qu'il s'agit des décimales de π, fait l'hypothèse que la série est issue de tirages aléatoires indépendants suivant une loi équirépartie. Il prend un risque de 10 % de rejeter cette hypothèse quand elle est vraie. Accepte-t-il cette hypothèse ?

☐ Oui ☐ Non ☐ Il ne peut pas conclure.

****

EXERCICE 4 **3 points**

Commun à tous les candidats

Partie A

On dispose d'un dé en forme de tétraèdre régulier, possédant une face bleue, deux faces rouges et une face verte ; on suppose le dé parfaitement équilibré.
Une partie consiste à effectuer deux lancers sucessifs et indépendants de ce dé. À chaque lancer on note la couleur de la face cachée.
On considère les évènements suivants :

E est l'évènement « à l'issue d'une partie, les deux faces notées sont vertes »,

F est l'évènement « à l'issue d une partie, les deux faces notées sont de la même couleur ».

1. Calculer les probabilités des évènements E et F ainsi que la probabilité de E sachant F.
2. On effectue dix parties identiques et indépendantes.
   Calculer la probabilité d'obtenir au moins deux fois l'évènement F au cours de ces dix parties (on en donnera une valeur approchée décimale à 10^{-3} près).

Partie B

On souhaite savoir si le dé utilisé peut être considéré comme parfaitement équilibré. Pour cela on numérote de 1 à 4 les quatre faces de ce dé, puis on lance, ce dé 160 fois en notant le nombre n_i de fois où chaque face est cachée ; on obtient les résultats suivants :

face i	1	2	3	4
effectif n_i	30	48	46	32

On note f_i la fréquence relative à la face n_i et d^2_{obs} le réel $\sum_{i=1}^{4}\left(f_i - \frac{1}{4}\right)^2$.
On simule ensuite 1 000 fois l'expérience consistant à tirer un chiffre au hasard 160 fois parmi l'ensemble (1 ; 2 ; 3 ; 4) puis, pour chaque simulation, on calcule
$d^2 = \sum_{i=1}^{4}\left(F_i - \frac{1}{4}\right)^2$, où F_i est la fréquence d'apparition du nombre i. Le 9$^{\text{e}}$ décile de la série statistique des 1 000 valeurs de d^2 est égal à 0,009 8.
Au vu de l'expérience réalisée et au risque de 10 %, peut-on considérer le dé comme parfaitement équilibré ?

****

EXERCICE 4 **5 points**

Commun à tous les candidats

1. Dans un stand de tir, un tireur effectue des tirs successifs pour atteindre un ballon afin de le crever. À chacun de ces tirs, il a la probabilité 0,2 de crever le ballon. Le tireur s'arrête quand le ballon est crevé. Les tirs successifs sont supposés indépendants.

a. Quelle est la probabilité qu'au bout de deux tirs le ballon soit intact ?

b. Quelle est la probabilité que deux tirs suffisent pour crever le ballon ?

c. Quelle est la probabilité p_n que n tirs suffisent pour crever le ballon ?

d. Pour quelles valeurs de n a-t-on $p_n > 0,99$?

2. Ce tireur participe au jeu suivant :

Dans un premier temps il lance un dé tétraédrique régulier dont les faces sont numérotées de 1 à 4 (la face obtenue avec un tel dé est la face cachée) ; soit k le numéro de la face obtenue. Le tireur se rend alors au stand de tir et il a droit à k tirs pour crever le ballon.

Démontrer que, si le dé est bien équilibré, la probabilité de crever le ballon est égale à 0,409 6 (on pourra utiliser un arbre pondéré).

3. Le tireur décide de tester le dé tétraédrique afin de savoir s'il est bien équilibré ou s'il est pipé. Pour cela il lance 200 fois ce dé et il obtient le tableau suivant :

Face k	1	2	3	4
Nombre de sorties de la face k	58	49	52	41

a. Calculer les fréquences de sorties f_k observées pour chacune des faces.

b. On pose $d^2 = \Sigma_{k=1}^{4}\left(f_k - \dfrac{1}{4}\right)^2$. Calculer d^2.

c. On effectue maintenant 1 000 simulations des 200 lancers d'un dé tétraédrique bien équilibré et on calcule pour chaque simulation le nombre d^2. On obtient pour la série statistique des 1 000 valeurs de d^2 les résultats suivants :

Minimum	D_1	Q_1	Médiane	Q_3	D_9	Maximum
0,001 24	0,001 92	0,002 35	0,002 81	0,003 45	0,004 52	0,010 15

Au risque de 10 %, peut-on considérer que ce dé est pipé 7

****

Annexe 5 : *3 sujets proposés lors de l'expérimentation d'une épreuve pratique (2008)*

Sujet 066 — Épreuve pratique de mathématiques — Fiche élève

Suite aléatoire

Énoncé

On considère une suite (S_n) définie par le lancer d'une pièce équilibrée de la façon suivante :

$$S_0 = 0 \quad \text{et} \quad \begin{cases} S_{n+1} = S_n + 1 & \text{si on obtient PILE} \\ S_{n+1} = S_n - 1 & \text{si on obtient FACE} \end{cases}.$$

On note A_n l'événement « obtenir $S_n = 0$ ».

On s'intéresse à la probabilité de réaliser l'événement A_n pour un entier n non nul donné.

Étude expérimentale

1. En utilisant un tableur, effectuer une simulation donnant les 11 premiers termes de 1 000 suites définies de la même façon que (S_n).

   Existe-t-il des valeurs de n pour lesquelles l'événement A_n est impossible ? Justifier votre réponse.

   Appeler l'examinateur pour présenter votre simulation et votre justification.

2. (a) Donner les fréquences d'apparition de l'événement A_n pour n variant de 1 à 10.

   (b) Faire d'autres simulations de même taille pour compléter le tableau suivant :

Fréquences d'apparition de A_n										
n	1	2	3	4	5	6	7	8	9	10
Simulation n° 1										
Simulation n° 2										
Simulation n° 3										
Simulation n° 4										
Simulation n° 5										

Appeler l'examinateur pour une vérification.

Étude mathématique

3. Déterminer les probabilités de réaliser les événements A_2, A_4 et A_6.

   Appeler l'examinateur pour une vérification.

4. Donner une expression de $p(A_n)$ en fonction de la parité de n.

Production demandée

- Présentation orale des premiers termes des suites puis du tableau des fréquences des 5 simulations ;
- Calcul de $p(A_2)$, de $p(A_4)$ et de $p(A_6)$;
- Justification de la méthode de calcul de $p(A_n)$.

---

1 / 1

Marche aléatoire

Énoncé

Un pion est placé sur la case de départ :

				Départ				

Le lancer d'une pièce bien équilibrée détermine le déplacement du pion.

- PILE, le pion se déplace vers la droite
- FACE, le pion se déplace vers la gauche

Un trajet est une succession de 4 déplacements. On s'intéresse à l'événement A : « le pion est revenu à la case départ après 4 déplacements ».

À chaque lancer, on associe le réel $+1$ si le résultat est PILE et -1 si le résultat est FACE.

Étude expérimentale

1. Simuler à l'aide du tableur de 200 à 2000 trajets du pion et estimer la fréquence de l'événement A. Compléter le tableau suivant :

Nombre d'essais	200	400	600	800	1000	1200	1400	1600	1800	2000
Fréquence de A										

Appeler l'examinateur pour vérifier le tableau obtenu.

Étude mathématique

2. On appelle X la variable aléatoire qui prend pour valeur la somme des quatre réels.
   (a) En précisant la méthode choisie, calculer les valeurs possibles de X et le nombre de trajets possibles.

   Appeler l'examinateur pour contrôler la réponse et lui indiquer la démarche prévue à la question suivante

   (b) Calculer la probabilité de l'événement A à l'aide d'un schéma de Bernoulli et comparer avec l'estimation obtenue.

Production demandée

- Réaliser une simulation en utilisant les fonctions appropriées.
- Donner une réponse argumentée à la question 2.

Étude d'un jeu

Énoncé

On lance trois dés bien équilibrés dont les six faces sont numérotées de 1 à 6.
Alice et Bob calculent la somme des trois nombres obtenus.
Si la somme obtenue est égale à 9, Alice gagne.
Si la somme obtenue est égale à 10, Bob gagne.
Dans tous les autres cas, la partie est annulée.

Le but de l'exercice est de déterminer qui, d'Alice ou de Bob, a la plus grande probabilité de gagner.

Étude expérimentale

1. Sur un tableur, réaliser une simulation de cette expérience aléatoire.

   Appeler l'examinateur pour valider cette simulation.

2. Sur un tableur, réaliser une simulation sur un échantillon de taille 1000 de cette expérience aléatoire et déterminer, pour cette simulation, les fréquences de réussite respectives d'Alice et de Bob.

   Appeler l'examinateur pour valider la feuille de calcul construite.

3. Est-il possible de conjecturer qui, d'Alice ou de Bob, a la plus grande probabilité de gagner ?

   Appeler l'examinateur pour lui fournir cette réponse et lui indiquer les méthodes prévues pour les démonstrations qui suivent.

Étude mathématique

On souhaite maintenant calculer la probabilité de gagner d'Alice et de Bob.

4. Répondre aux deux questions suivantes (dans n'importe quel ordre) :
   – Calculer la probabilité de gagner d'Alice et de Bob.
   – Qui, d'Alice ou de Bob, a la plus grande probabilité de gagner ?

Production demandée

– Bilan de la simulation de la question 2 ;
– Réponse orale à la question 3 ;
– Réponses argumentées à la question 4.

Annexe 6 : programme de seconde (« programmes 2010 »)

3. Statistiques et probabilités

Pour des questions de présentation du programme, les cadres relatifs à l'enseignement des statistiques et des probabilit sont présentés séparément à la suite l'un de l'autre. Pour autant, ces enseignements sont en relation étroite l'un avec l'au et doivent faire l'objet d'allers et retours.

Objectifs visés par l'enseignement des statistiques et probabilités à l'occasion de résolutions de problèmes

dans le cadre de l'analyse de données, rendre les élèves capables

- de déterminer et interpréter des résumés d'une série statistique ;
- de réaliser la comparaison de deux séries statistiques à l'aide d'indicateurs de position et de dispersion, ou de la cour des fréquences cumulées ;

dans le cadre de l'échantillonnage

- faire réfléchir les élèves à la conception et la mise en œuvre d'une simulation ;
- sensibiliser les élèves à la fluctuation d'échantillonnage, aux notions d'intervalle de fluctuation et d'intervalle de confia et à l'utilisation qui peut en être faite.

CONTENUS	CAPACITÉS ATTENDUES	COMMENTAIRES
Statistique descriptive, analyse de données Caractéristiques de position et de dispersion • médiane, quartiles ; • moyenne.	• Utiliser un logiciel (par exemple, un tableur) ou une calculatrice pour étudier une série statistique. • Passer des effectifs aux fréquences, calculer les caractéristiques d'une série définie par effectifs ou fréquences. • Calculer des effectifs cumulés, des fréquences cumulées. • Représenter une série statistique graphiquement (nuage de points, histogramme, courbe des fréquences cumulées).	L'objectif est de faire réfléchir les élèves sur des données réelles, riches et variées (issues, par exemple, d'un fichier mis à disposition par l'INSEE), synthétiser l'information et proposer des représentations pertinentes.
Échantillonnage Notion d'échantillon. Intervalle de fluctuation d'une fréquence au seuil de 95%*. Réalisation d'une simulation.	• Concevoir, mettre en œuvre et exploiter des simulations de situations concrètes à l'aide du tableur ou d'une calculatrice. • Exploiter et faire une analyse critique d'un résultat d'échantillonnage.	Un échantillon de taille n est constitué des résultats de n répétitions indépendantes de la même expérience. À l'occasion de la mise en place d'une simulation, on peut : • utiliser les fonctions logiques d'un tableur ou d'une calculatrice, ◇ mettre en place des instructions conditionnelles dans un algorithme. L'objectif est d'amener les élèves à un questionnement lors des activités suivantes : • l'estimation d'une proportion inconnue à partir d'un échantillon ; • la prise de décision à partir d'un échantillon.

* L'intervalle de fluctuation au seuil de 95%, relatif aux échantillons de taille n, est l'intervalle centré autour de p, proportion du caractère dans la population, où se situe, avec une probabilité égale à 0,95, la fréquence observée dans un échantillon de taille n. Cet intervalle peut être obtenu, de façon approchée, par simulation. Le professeur peut indiquer aux élèves le résultat suivant, utilisable dans la pratique pour des échantillons de taille $n \geqslant 25$ et des proportions p du caractère comprises entre 0,2 et 0,8 : si f désigne la fréquence du caractère dans l'échantillon, f appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$ avec une probabilité d'au moins 0,95. Le professeur peut faire percevoir expérimentalement la validité de cette propriété mais **elle n'est pas exigible**.

Objectifs visés par l'enseignement des statistiques et probabilités à l'occasion de résolutions de problèmes

dans le cadre des probabilités, rendre les élèves capables :

- d'étudier et modéliser des expériences relevant de l'équiprobabilité (par exemple, lancers de pièces ou de dés, tirage d cartes) ;
- de proposer un modèle probabiliste à partir de l'observation de fréquences dans des situations simples ;
- d'interpréter des événements de manière ensembliste ;
- de mener à bien des calculs de probabilité.

Les situations étudiées concernent des expériences à une ou plusieurs épreuves.

◇ La répétition d'expériences aléatoires peut donner lieu à l'écriture d'algorithmes (marches aléatoires).

CONTENUS	CAPACITÉS ATTENDUES	COMMENTAIRES
Probabilité sur un ensemble fini		
Probabilité d'un événement.	• Déterminer la probabilité d'événements dans des situations d'équiprobabilité. • Utiliser des modèles définis à partir de fréquences observées.	La probabilité d'un événement est définie comme la somme des probabilités des événements élémentaires qui le constituent. Pour les calculs de probabilités, on utilise des arbres, des diagrammes ou des tableaux.
Réunion et intersection de deux événements, formule : $p(A \cup B) + p(A \cap B) = p(A) + p(B)$.	• Connaître et exploiter cette formule.	

****

Annexes du chapitre 11

Annexe 7 : ressources pour la classe de seconde, algorithmique (extrait)

Nous reproduisons intégralement la partie « Algorithmes et probabilités ».

Algorithmes et probabilités

Les algorithmes proposés ci-après s'insèrent dans le cadre de la simulation, et par conséquent de l'approche dite « fréquentiste » des probabilités.

1 / Le jeu du lièvre et de la tortue

Règle du jeu.

À chaque tour, on lance un dé. Si le 6 sort, alors le lièvre gagne la partie, sinon la tortue avance d'une case. La tortue gagne quand elle a avancé 6 fois.

Question : le jeu est-il à l'avantage du lièvre ou de la tortue ?

a. Algorithme 1 : Simulation d'une partie sans boucle

La partie se finit en au plus six lancés. Il est donc possible de simuler une partie sans avoir recours à une boucle.

```
Variables
    dé : la face du dé tirée au hasard
    tour : compte le nombre de tours que dure la partie
Initialisation
    dé prend une valeur entière aléatoire entre 1 et 6 compris
    tour prend la valeur 1
Traitement
    Si dé < 6 alors
        dé prend une valeur entière aléatoire entre 1 et 6 compris
        tour augmente de 1
    Si dé < 6 alors
        dé prend une valeur entière aléatoire entre 1 et 6 compris
        tour augmente de 1
    Si dé < 6 alors
        dé prend une valeur entière aléatoire entre 1 et 6 compris
        tour augmente de 1
    Si dé < 6 alors
        dé prend une valeur entière aléatoire entre 1 et 6 compris
        tour augmente de 1
    Si dé < 6 alors
        dé prend une valeur entière aléatoire entre 1 et 6 compris
        tour augmente de 1
Sortie
    Si dé = 6 alors
        Affiche « Le lièvre gagne »
    sinon
        Affiche « La tortue gagne »
    Affiche tour
```

Remarques :

– Si le support de programmation sur lequel est transposé l'algorithme ne dispose pas des fonctionnalités de recopie de texte, la mise en place d'une boucle peut se justifier.

– Il est possible de modifier un peu la modélisation du jeu, afin de simplifier sa mise en œuvre, en particulier sur tableur ; en effet, la partie est équivalente aux lancers de six dés. Le lièvre gagne s'il existe au moins un six parmi les résultats.

– De cette façon un seul test peut suffire, à condition de disposer de l'instruction appropriée (comme l'instruction NB.SI du tableur).

– Cependant, l'arithmétique booléenne peut aussi se modéliser avec des sommes ou des produits d'entiers. Par exemple, le produit des six nombres (6 – *dé*) vaut 0 si et seulement il y a au moins un six.

b. Algorithme 2 : Cumuler un grand nombre d'expériences

Il suffit d'aménager le programme afin d'insérer la simulation d'une partie dans une boucle et de compter le nombre de parties gagnées par le lièvre ou la tortue. L'intérêt d'un langage de programmation devient évident : l'itération est très rapide aussi bien à écrire, modifier qu'à exécuter (ce qui n'est pas le cas avec le tableur). On pourra noter, à cette occasion, que certains langages sont beaucoup plus rapides que d'autres.

```
Variables
    dé : la face du dé tirée au hasard
    N : le nombre de parties à simuler
    k : le compteur de boucle
    tortue : le nombre de parties gagnées par la tortue
Initialisation
    tortue prend une valeur 0
Traitement
    Pour k de 1 à N
        dé prend une valeur entière aléatoire entre 1 et 6 compris
        Si dé < 6 alors
            dé prend une valeur entière aléatoire entre 1 et 6 compris
        Si dé < 6 alors
            dé prend une valeur entière aléatoire entre 1 et 6 compris
        Si dé < 6 alors
            dé prend une valeur entière aléatoire entre 1 et 6 compris
        Si dé < 6 alors
            dé prend une valeur entière aléatoire entre 1 et 6 compris
        Si dé < 6 alors
            dé prend une valeur entière aléatoire entre 1 et 6 compris
        Si dé < 6 alors
            tortue prend la valeur tortue + 1
Sortie
    Affiche tortue .
```

Traduction XCAS

```
N:=10;
Tortue:=0;
for (K:=1;K<=N;K:=K+1){
  dé:=rand(6)+1;
  if (dé<6) {dé:=rand(6)+1;}
  if (dé<6) {dé:=rand(6)+1;}
  if (dé<6) {dé:=rand(6)+1;}
  if (dé<6) {dé:=rand(6)+1;}
  if (dé<6) {dé:=rand(6)+1;}
  if (dé<6) {Tortue:=Tortue+1;}
};
print (Tortue);
```

c. Algorithme 3 : Avec une structure itérative conditionnelle.

Évidemment, plutôt que de répéter 6 fois les mêmes instructions, il est possible de simuler une partie à l'aide d'une boucle. De cette façon, il sera facile d'expérimenter de nouveaux jeux en modifiant le nombre de cases que doit parcourir la tortue

Variables

    dé : la face du dé tirée au hasard
    case : le numéro de la case sur laquelle se trouve la tortue
    N : le nombre de cases que doit parcourir la tortue pour gagner.

Initialisation

    N **prend la valeur** 6
    case **prend la valeur 0.**

Traitement

    Répète
        dé **prend une valeur** entière aléatoire entre 1 et 6 inclus.
        Si dé < 6 **alors**
            case **prend la valeur** case + 1
    jusqu'à [dé = 6 ou case = N **]**

Sortie

    Si case = N **alors**
        Affiche « La tortue gagne »
    Sinon
        Affiche « Le lièvre gagne »

Traduction SCILAB

```
N=6;
Ncase=0;
de=0;
while (de<6 | Ncase<N) do
  de=floor(rand()*6+1);
  if (de<6)
    Ncase=Ncase+1;
  end;
end;
if (Ncase==6)
  disp("La tortue gagne");
else
  disp("Le lièvre gagne");
end;
```

Remarques :

« *case* » est un mot-clé du langage SCILAB ; la variable s'appelle donc « *Ncase* ».

La structure *repeat..until* n'existe pas dans SCILAB, le code est donc légèrement aménagé par rapport à l'algorithme. Pour entrer dans la boucle une première fois, la variable « *de* » est initialisée avec la valeur arbitraire 0.

2 / Coïncidence de date d'anniversaire dans une classe

Dans la vie courante certaines coïncidences apparaissent « extraordinaires » (comme rencontrer par hasard quelqu'un de connu à des centaines de kilomètres de chez soi). Malheureusement bien souvent ces coïncidences ne se prêtent pas facilement à une modélisation qui permettrait un calcul de probabilité ou une simulation.

Le problème évoqué dans ce paragraphe ne pose pas de grandes difficultés de modélisation; pour autant, le résultat s'avèrera sans doute étonnant pour de nombreux élèves.

Sa mise en place algorithmique peut être l'occasion de travailler des questions proches de celles des tris qui font souvent intervenir deux boucles imbriquées.

Quelle est la probabilité que dans une classe de 30 élèves, il y ait au moins deux élèves qui partagent la même date d'anniversaire ?

Pour effectuer une simulation, il s'agit dans un premier temps de tirer les 30 dates d'anniversaires au sort (parmi 365 jours, en supposant les dates d'anniversaire uniformément réparties sur l'année civile); il faudra ensuite chercher si deux dates coïncident.

Les dates sont mémorisées dans un tableau. Comme c'est souvent l'usage, on note entre crochets l'indice du tableau. Dans l'exemple, on considère que les indices du tableau commencent à 0.

Algorithme 4

Variables
    dates : tableau des trente jours d'anniversaire
    trouvé : un booléen qui indique si deux dates coïncident.
    k, p: deux compteurs de boucles.
Initialisation
    Pour k **de** 0 **à** 29
        dates[k] **prend une valeur** entière aléatoire comprise entre 1 et 365 inclus

    trouvé **prend la valeur** faux
Traitement
    Pour k **de** 0 **à** 28
        Pour p **de** k+1 **à** 29
            Si dates[k] = dates[p] **alors**
                trouvé **prend la valeur vrai**
Sortie
    Affiche trouvé

Si le tirage au sort des dates se fait aisément sur tableur, il n'en va pas de même de la recherche de dates identiques (du fait des deux boucles).

Traduction SCILAB :

```
dates=floor(rand(30,1)*365+1);
trouve=%F;
for k=1:29
  for p=k+1:30
    if (dates(p)==dates(k))
      trouve=%T;
    end;
  end;
end;
if (trouve)
  disp("Deux personnes ont même
     anniversaire");
else
  disp("Pas deux anniversaires
     communs");
end;
```

Traduction SCILAB (1000 expériences) :

```
N=0;
for i=1:1000
  dates=floor(rand(30,1)*365+1);
  trouve=%F;
  for k=1:29
    for p=k+1:30
      if (dates(p)==dates(k))
        trouve=%T;
      end;
    end;
  end;
  if (trouve)
    N=N+1;
  end;
end;
disp("Il y a "+string(N)+" coïncidences");
```

****

Annexe 8 : ressources pour la classe de première : statistique et probabilités (extraits)

Ce document comporte 76 pages, nous ne pouvons l'intégrer dans sa totalité. Nous présentons ici le sommaire, et quelques unes des pages évoquées dans la thèse.

SOMMAIRE

pages 23 à 25 :

❖ **Le paradoxe de Saint-Pétersbourg**

Formulé par Nicolas Bernoulli en 1713, ce problème a été approfondi par son cousin Daniel Bernoulli dans l'ouvrage *Les transactions de l'Académie de Saint-Pétersbourg*, ce qui lui a valu son nom.

Énoncé

Un joueur joue contre la banque au jeu de « pile ou face », en misant toujours sur « face ». Il adopte la stratégie suivante : il mise un euro au premier coup, et s'il perd, double la mise au coup suivant, tant que « face » ne sort pas. S'il gagne, il récupère sa mise augmentée d'une somme équivalente à cette mise. Le joueur dispose d'une fortune limitée, qui lui permet de perdre au maximum n coups consécutifs et, si « pile » sort n fois de suite, le joueur ne peut plus miser et arrête le jeu. La fortune de la banque, elle, n'est pas limitée.

Une partie consiste pour le joueur à jouer, si sa fortune le lui permet, jusqu'à ce que « face » sorte.

Il s'agit de déterminer la probabilité qu'a le joueur de gagner une partie, son gain algébrique moyen par partie, et d'analyser l'intérêt pour le joueur de jouer à ce jeu.

Traitement mathématique

Pour modéliser la situation, on suppose que le joueur lance la pièce n fois : si « face » sort avant le n-ième coup, le joueur ne mise rien les coups suivants. Lorsqu'il joue n fois de suite à « pile ou face », on note :

– A_n l'événement « le joueur obtient n piles » ; $G=\overline{A_n}$ l'événement : « le joueur gagne la partie » ;
– X la variable aléatoire qui comptabilise le rang de la première face, et l'on convient que ce rang est égal à 0 si « face » ne sort pas ;
– Y la variable aléatoire qui donne le gain algébrique du joueur.

On envisage d'abord le cas où le joueur dispose d'une fortune limitée, par exemple à 1000 €.

Le joueur double sa mise tant qu'il perd. Sa fortune lui permet de tenir n coups, où il mise successivement (en euro) 1, 2, 2^2, ..., 2^{n-1}, tant que $1+2+...+2^{n-1}\leq 1000$. La formule sommatoire sur les suites géométriques simplifie cette inégalité en : $2^n-1\leq 1000$. D'où $n=9$.

On obtient $P(A_9)=\dfrac{1}{2^9}\approx 0{,}002$, d'où $P(G)=1-P(A_9)=1-\dfrac{1}{2^9}\approx 0{,}998$.

La variable aléatoire X suit la loi géométrique tronquée de paramètres 9 et $\dfrac{1}{2}$. Elle prend les valeurs entières de 0 à 9, avec $P(X=0)=P(A_9)=\dfrac{1}{2^9}$ et, pour k compris entre 1 et 9 : $P(X=k)=\dfrac{1}{2^{k-1}}\left(1-\dfrac{1}{2}\right)=\dfrac{1}{2^k}$.

On vérifie bien que $\sum_{k=0}^{9} P(X=k) = \left(\frac{1}{2}+\frac{1}{2^2}+\ldots+\frac{1}{2^9}\right)+\frac{1}{2^9}=1$.

Déterminons les valeurs de *Y*.

Si « face » sort pour la première fois au *k*-ième coup (avec $1 \leq k \leq n$), le joueur a misé au total une somme en euro égale à $1+2+\ldots+2^{k-1}$, il gagne le double de sa dernière mise, soit $2 \times 2^{k-1}$. Son gain algébrique est donc égal à $2^k-(1+2+\ldots+2^{k-1})$, c'est-à-dire 1 €.

Si « face » ne sort pas, le joueur a perdu toutes ses mises, soit (en euro) $1+2+\ldots+2^8=2^9-1=511$.

On en tire la loi de la variable aléatoire *Y* et son espérance mathématique :

Valeurs de *Y*	$+1$	$-(2^9-1)$
Probabilités	$1-\frac{1}{2^9}$	$\frac{1}{2^9}$

$$E(Y)=1\times\left(1-\frac{1}{2^9}\right)-(2^9-1)\times\frac{1}{2^9}=0.$$

<u>Simulation de 1000 parties en 9 coups au plus sur un tableur</u>

On code la sortie de « face » par « 1 », celle de « pile » par « 0 ».

On place en A1 la formule $\boxed{\text{=ENT(2*ALEA())}}$,

puis en B1 la formule $\boxed{\text{=SI(OU(A1=1;A1="");"";ENT(2*ALEA()))}}$, que l'on recopie jusqu'en I1 ;

enfin on place en K1 la formule $\boxed{\text{=SI(SOMME(A1:I1)=0;"PERDU";"GAGNE")}}$.

Les formules précédentes sont recopiées jusqu'à la ligne 1000.

Il reste alors en décompter en M1 le nombre de parties perdues, avec la formule :

$\boxed{\text{=NB.SI(K1:K1000;"PERDU")}}$.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	1										GAGNE	Nombre de parties perdues	6	
2	1										GAGNE			
3	0	0	0	1							GAGNE			
4	0	1									GAGNE			
5	1										GAGNE			
6	0	1									GAGNE			
7	1										GAGNE			
8	0	1									GAGNE			
9	1										GAGNE			
10	1										GAGNE			
11	1										GAGNE			
12	0	0	0	0	0	1					GAGNE			
13	0	0	0	0	1						GAGNE			
14	1										GAGNE			
15	0	0	0	0	0	0	0	0	0		PERDU			
16	0	1									GAGNE			
17	1										GAGNE			
18	0	1									GAGNE			

Quoique faible, la probabilité de perdre n'est pas négligeable. Sur la simulation précédente, on s'aperçoit que le joueur perd effectivement 6 parties sur 1000. Il perd donc six fois 511 €, soit 3066 €. Il a gagné 994 parties qui lui rapportent chacune 1 €, soit un gain total de 994 €. Il a donc perdu 2972 euros sur 1000 parties, soit environ 3 euros par partie en moyenne.

Conclusions de l'étude : deux paradoxes

Chaque partie gagnée rapporte 1 € au joueur. Si sa fortune était illimitée (ou simplement très grande), la probabilité de gagner, égale à $1-\frac{1}{2^n}$, aurait pour limite 1, et permettrait au joueur de gagner chaque partie. Il semble donc que la stratégie du joueur constitue une « martingale » infaillible. Le jeu semble favorable au joueur.

Cependant, puisque $E(Y)=0$, le jeu est honnête. La stratégie mise en place donne une espérance de gain identique à celle du simple jeu de pile ou face. C'est un premier paradoxe.

Par ailleurs, ce problème montre la limite de la notion d'espérance pour juger si un jeu est favorable. En effet, la simulation précédente a révélé que la perte est importante, et qu'elle se produit plusieurs fois sur 1000 parties. Peu de joueurs s'aventureraient dans un jeu pourtant honnête où l'on risque de perdre gros, même si ce risque est faible, alors que l'on gagne peu. C'est là le deuxième paradoxe.

(…)

page 35 :

2. Le quorum

Une association comprenant 30 adhérents organise chaque année une assemblée générale. Les statistiques montrent que chaque adhérent assiste à l'assemblée avec la probabilité 80 %. Les décisions prises par l'assemblée n'ont de valeur légale que lorsque plus de la moitié des adhérents assiste à l'assemblée.

Quelle est la probabilité que, lors de la prochaine assemblée, le quorum soit atteint ?

****

Annexe 9 : programmes de première S (« programmes 2010 »)

B.O. Bulletin officiel spécial n° 9 du 30 septembre 2010

3. Statistiques et probabilités

L'étude et la comparaison de séries statistiques menées en classe de seconde se poursuivent avec la mise en place de nouveaux outils dans l'analyse de données. L'objectif est de faire réfléchir les élèves sur des données réelles, riches et variées (issues, par exemple, de fichiers mis à disposition par l'Insee).
La notion de loi de probabilité d'une variable aléatoire permet de modéliser des situations aléatoires, d'en proposer un traitement probabiliste et de justifier certains faits observés expérimentalement en classe de seconde.
L'utilisation des arbres pondérés est développée pour modéliser la répétition d'expériences identiques et indépendantes. Elle est restreinte à ce cadre afin d'éviter toute confusion avec des situations relevant des probabilités conditionnelles.
Dans le cas particulier d'expériences identiques et indépendantes à deux issues, on introduit la loi binomiale. En s'appuyant sur cette loi, on poursuit la formation des élèves dans le domaine de l'échantillonnage.

CONTENUS	CAPACITÉS ATTENDUES	COMMENTAIRES
Statistique descriptive, analyse de données Caractéristiques de dispersion : variance, écart-type. Diagramme en boîte.	• Utiliser de façon appropriée les deux couples usuels qui permettent de résumer une série statistique : (moyenne, écart-type) et (médiane, écart interquartile). • Étudier une série statistique ou mener une comparaison pertinente de deux séries statistiques à l'aide d'un logiciel ou d'une calculatrice.	On utilise la calculatrice ou un logiciel pour déterminer la variance et l'écart-type d'une série statistique. Des travaux réalisés à l'aide d'un logiciel permettent de faire observer des exemples d'effets de structure lors du calcul de moyennes.
Probabilités Variable aléatoire discrète et loi de probabilité. Espérance, variance et écart-type.	• Déterminer et exploiter la loi d'une variable aléatoire. • Interpréter l'espérance comme valeur moyenne dans le cas d'un grand nombre de répétitions.	À l'aide de simulations et d'une approche heuristique de la loi des grands nombres, on fait le lien avec la moyenne et la variance d'une série de données. On exploite les fonctionnalités de la calculatrice ou d'un logiciel pour déterminer l'espérance, la variance et l'écart-type d'une variable aléatoire. ⊟ On démontre les formules suivantes sur l'espérance et la variance : $E(aX+b)=aE(X)+b$ et $V(aX)=a^2V(X)$.

Modèle de la répétition d'expériences identiques et indépendantes à deux ou trois issues.	• Représenter la répétition d'expériences identiques et indépendantes par un arbre pondéré. • Utiliser cette représentation pour déterminer la loi d'une variable aléatoire associée à une telle situation.	Pour la répétition d'expériences identiques et indépendantes, la probabilité d'une liste de résultats est le produit des probabilités de chaque résultat. La notion de probabilité conditionnelle est hors programme. On peut aussi traiter quelques situations autour de la loi géométrique tronquée. ◇ On peut simuler la loi géométrique tronquée avec un algorithme.
Épreuve de Bernoulli, loi de Bernoulli. Schéma de Bernoulli, loi binomiale (loi du nombre de succès). Coefficients binomiaux, triangle de Pascal.	• Reconnaître des situations relevant de la loi binomiale. • Calculer une probabilité dans le cadre de la loi binomiale.	La représentation à l'aide d'un arbre est privilégiée : il s'agit ici d'installer une représentation mentale efficace. On peut ainsi : faciliter la découverte de la loi binomiale pour des petites valeurs de n ($n \leq 4$) ; introduire le coefficient binomial $\binom{n}{k}$ comme nombre de chemins de l'arbre réalisant k succès pour n répétitions ; établir enfin la formule générale de la loi binomiale.
	▣ Démontrer que $\binom{n}{k}+\binom{n}{k+1}=\binom{n+1}{k+1}$. • Représenter graphiquement la loi binomiale.	Cette égalité est établie en raisonnant sur le nombre de chemins réalisant $k+1$ succès pour $n+1$ répétitions. On établit également la propriété de symétrie des coefficients binomiaux. L'utilisation des coefficients binomiaux dans des problèmes de dénombrement et leur écriture à l'aide des factorielles ne sont pas des attendus du programme. En pratique, on utilise une calculatrice ou un logiciel pour obtenir les valeurs des coefficients binomiaux, calculer directement des probabilités et représenter graphiquement la loi binomiale.
Espérance, variance et écart-type de la loi binomiale.	• Utiliser l'espérance d'une loi binomiale dans des contextes variés.	La formule donnant l'espérance de la loi binomiale est conjecturée puis admise, celle de la variance est admise. ◇ On peut simuler la loi binomiale avec un algorithme.
Échantillonnage Utilisation de la loi binomiale pour une prise de décision à partir d'une fréquence.	• Exploiter l'intervalle de fluctuation à un seuil donné, déterminé à l'aide de la loi binomiale, pour rejeter ou non une hypothèse sur une proportion.	L'objectif est d'amener les élèves à expérimenter la notion de « différence significative » par rapport à une valeur attendue et à remarquer que, pour une taille de l'échantillon importante, on conforte les résultats vus en classe de seconde. ◇ L'intervalle de fluctuation peut être déterminé à l'aide d'un tableur ou d'un algorithme. Le vocabulaire des tests (test d'hypothèse, hypothèse nulle, risque de première espèce) est hors programme.

Annexes du chapitre 13

Annexe 10 : Note de l'Inspection Générale de mathématiques aux concepteurs de sujets

Les épreuves écrites au baccalauréat et leur évaluation

Ce texte s'inscrit dans le prolongement des notes de service définissant les modalités des épreuves écrites de mathématiques des différentes séries de baccalauréat.

Les principes

Actuellement, les épreuves écrites couvrent assez bien les différents champs des programmes mais ne prennent pas suffisamment en compte les compétences qu'ils visent à développer.
C'est pourquoi il est souhaitable de :

- poursuivre l'évolution progressive du contenu des épreuves, déjà engagée avec l'introduction de QCM, de ROC, de questions ouvertes ;
- faire évoluer conjointement l'évaluation des copies en essayant de les apprécier en relation avec les compétences attendues des candidats.

Ces évolutions induisent plusieurs exigences :

- mettre en phase le contenu des sujets avec les objectifs des programmes, conformément à ce que prévoit la note de service définissant l'épreuve écrite ;
- identifier *a priori* des compétences à évaluer et bâtir les sujets en fonction de ces compétences ;
- clarifier pour les élèves le contrat d'évaluation dans la formulation des questions leur donnant une plus grande initiative, en les incitant notamment à laisser des traces de leurs recherches ;
- organiser la correction en tenant compte de ces règles.

Les compétences à évaluer

Certaines compétences, que l'on peut qualifier de « compétences de base », sont évaluées dans toutes les épreuves de baccalauréat ; il s'agit de la mobilisation et de la restitution de connaissances et de la capacité à appliquer des méthodes. Ces compétences sont présentes à de nombreuses occasions dans tous les exercices d'un même sujet. Il n'est pas utile de les repérer précisément.

♦ Les sujets des baccalauréats des séries S, ES et L spécialité mathématiques **doivent permettre, de plus, d'évaluer** la maîtrise de compétences évoluées parmi les suivantes :
  - prendre des initiatives, choisir un modèle, émettre une conjecture, expérimenter ;
  - raisonner, démontrer, élaborer une démarche ;
  - évaluer, critiquer un résultat, vérifier la validité d'un résultat ou d'une méthode.

♦ Les sujets des baccalauréats technologiques et du baccalauréat L (épreuve anticipée mathématiques-informatique) **doivent permettre d'évaluer** les deux compétences évoluées suivantes :
  - montrer une certaine autonomie dans le traitement de l'information (rechercher, organiser, traiter l'information) ;
  - développer une démarche connue, mettre en forme un raisonnement.

L'élaboration des sujets

L'élaboration des sujets suit les recommandations suivantes :

- la difficulté globale des sujets n'est pas changée ;
- la nature des exercices et la façon de poser les questions permettent d'apprécier, au moins une fois dans le sujet, l'acquisition des compétences évoluées correspondant à la série ;
- pour une question qui évalue les compétences évoluées mentionnées ci-dessus, un texte indique clairement au candidat qu'on attend qu'il laisse des traces de ses démarches, même non abouties.

Les consignes de correction et les travaux des commissions d'entente

Les consignes de correction précisent la compétence évoluée principalement visée par la ou les questions correspondantes ou par l'ensemble de l'exercice ; elles donnent les attendus et les savoir faire évalués dans chaque question ou chaque exercice évaluant une compétence évoluée.
La correction se fait avec un barème de notation finalisé par la commission d'entente.

Annexe

Exemple de texte porté dans l'en-tête du sujet :

> « Le candidat est invité à faire figurer sur la copie toute trace de recherche, même incomplète ou non fructueuse, qu'il aura développée. Il est rappelé que la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies. »

Exemple de texte inséré avant une question nécessitant une prise d'initiative :

> « Dans cette question [dans cet exercice], toute trace de recherche, même incomplète, ou d'initiative même non fructueuse, sera prise en compte dans l'évaluation. »

Annexes du chapitre 17

Annexe 11 : tableau des valeurs des χ^2 observés et interprétation

Nous avons fait le test du χ^2 pour étudier si les écarts observés dans les distributions des réponses, selon les modalités ou selon les niveaux, sont significatifs.

NS : non significatif.

	tableau 7	tableau 8	tableau 9	tableau 11	tableau 12
χ^2	2,955	4,719	0,124	2,980	5
probabilité	0,399	0,193	0,725	0,561	0,287
seuil au risque 5 %	7,815	7,815	3,841	9,488	9,488
conclusion	NS	NS	NS	NS	NS

Annexes du chapitre 18

<u>Annexe 12 : transcription des réponses des élèves à la question B.</u>

Les élèves qui n'ont pas répondu à cette question n'apparaissent pas dans le tableau.
Dans la première colonne apparaissent, pour un élève, son niveau, l'énoncé reçu et sa référence :

T = terminale et P = première,

A = énoncé 1 et B = énoncé 2

La référence est un numéro de 1 à 157.

Les réponses des élèves de la classe qui n'a pas été observée dans les mêmes conditions que les autres classes (cf. chapitre 15) correspondent aux références de 45 à 56, et ne sont pas prises en compte dans notre étude.

Nous avons essayé de faire une transcription aussi fidèle que possible des réponses.

<table>
<tr><td>TB1</td><td>Pour connaître le nombre de cases séparant la case de départ de la case d'arrivée, il faudrait simuler un grand nombre de fois la simulation de la question A1 et puis faire la moyenne des sommes obtenues à la case D1.</td></tr>
<tr><td>TB4</td><td>Je conseille à Monsieur Dupond d'effectuer la simulation en A1 plusieurs fois, puis de faire la moyenne des cases D22.
Cette moyenne correspond au nombre de cases à choisir pour un jeu d'environ 30 minutes.</td></tr>
<tr><td>TB5</td><td>Pour conseiller Mr. Dupond, on calcule l'espérance d'un lancé.
<table>
<tr><td>X=0</td><td>4</td><td>5</td><td>6</td><td>7</td><td></td><td>avec toutes les valeurs de X possibles</td></tr>
<tr><td>p= $\frac{4}{16}$</td><td>$\frac{2}{16}$</td><td>$\frac{4}{16}$</td><td>$\frac{2}{16}$</td><td>$\frac{2}{16}$</td><td></td><td></td></tr>
</table>
$0 \times \frac{4}{16} + 4 \times \frac{2}{16} + 5 \times \frac{4}{16} + 6 \times \frac{2}{16} + 7 \times \frac{2}{16}$ = 3,375. L'espérance est donc 3,375.
$\frac{20}{3,75}$ = 5,9. Un joueur finit donc en moyenne, une partie de 20 tours de 30 minutes en ~~6 minutes.~~ avançant de 6 cases. Je lui conseillerai donc de jouer avec 6 cases.</td></tr>
<tr><td>TB6</td><td>En 10 tours, il serait en moyenne arrivé à 10 cases → $10 \times \frac{4}{10} \times 0 + 10\times\frac{3}{10} \times 1 + 10\times\frac{2}{10} \times 2 + 10\times\frac{3}{10} \times 3$ = 10.
Par conséquent en 20 tours, il effecturait 20 cases.
On conseillera à M. Dupond de construire un plateau de 20 cases.</td></tr>
<tr><td>TB7</td><td>• p(X=1) est la plus importante. p(X=1) = $\frac{6}{16}$
• p(X=0) = p(X=2)</td></tr>
</table>

	Le joueur a $\frac{1}{4}$ de chances de faire d'avancer de 2 cases d'un coup et $\frac{1}{4}$ de chance de ne pas avancer du tout, ce qui ferait une moyenne d'un avancement de 1 case sur deux tours. • p(X=3) = $\frac{2}{16}$ Il est donc peu probable d'avancer de 3 cases d'un coup. • Si jamais le joueur ne fait que des lancés l'ammenant à avancer de 3 cases à la fois, il faudrait 20 ×3 = 60 cases. • Si jamais le joueur ne fait que des lancés l'ammenant à avancer d'1 case à la fois, il faudrait 20 cases. Mais il y a aussi plus de probabilité d'arriver à avancer une fois de 0 cases et une autre fois de 2 cases p(0+2)=$\frac{1}{2}$. Donc d'avancer de 1 case sur 2 tours. Soit 10 cases au bout de 20 tours Il faudrait donc une quinzaine de cases.
TB8	Je lui conseillerai de faire la moyenne des cases avancées par tour en tenant compte des probabilités. c'est à dire p(0) ×0 + p(1) ×1 + p(2) ×2 + p(3) ×3 Puis de mettre le nombre de cases dont l'on avance en moyenne sur 20 tours. C'est à dire de multiplier le résultat du calcul ci-dessus par 20.
TB9	On sait que le joueur doit gagner à l'issue de 20 tours. Il va donc lancer 20 fois les 2 dés. On pourrait conseiller à Monsieur Dupond de simuler plusieurs parties afin de prendre en considération le nombre de cases parcourues en 20 tours. (il regarde donc la cellule D1 du tableur) et de calculer la probabilité des diverses propositions. Ainsi il choisira de mettre le nombre de cases dont la probabilité observée aura été la plus grande.

TB10	Si je pouvais aider M. Dupond à déterminer le nombre de cases nécessaires pour son jeu, je lui conseillerai que, puisqu'il veut qu'une partie ne dure surtout pas plus de 20 tours, il observe les probabilités ci-dessus et qu'il constate, que la probabilité la plus forte (mis à part le cas des doubles où le joueur n'avancerait pas) est pour le cas ou le joueur X n'avance que d'une seule case par tour. Soit 1×20 = 20 cases en tout, soit 18 cases entre la première et la dernière.
TB11	Nous pourrions conseiller à Mr Dupond de ne prendre qu'en considération le fait qu'un joueur tombe majoritairement sur des dés ayant un écart de 1. (puisque p(0) et p(2) sont égales mais lorsque le joueur obtient 0, n'avance pas du tout alors qu'avec 2, il avance 2 fois plus qu'en 1)
TB12	On peut lui conseiller de simuler 1000 fois une partie de 20 tours et de faire la moyenne du nombre de cases parcourues en une partie.
TB13	La plus grande probabilité est X=1, donc le joueur a plus de chance de se déplacer d'une case que de 0, 2 ou 3. Monsieur Dupond devrait donc prévoir que le joueur se déplace d'une case à chaque coup, soit en 20 coups de 20 cases. Je lui conseille donc de mettre 20 cases entre le départ et l'arrivée.
TB14	Il y a une grande probabilités que les joueurs parcourent 1 case par tour on peut donc penser qu'il faudra
TB15	$E = \sum p_i x_i = 0 \times \frac{1}{4} + 1 \times \frac{3}{8} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} = 1,5$ L'espérance est de 1,5 par tours

	En multipliant ce nombre par le nombre total de tour (20), on trouve alors $n_{\text{cases pour finir le jeu en 30 min}} = 30$ Monsieur Dupond devrait placer 30 cases entre le départ et l'arrivée pour que le jeu se termine <u>en moyenne</u> en 30 minutes.
TB16	Un joueur a une plus grande probabilité d'avancer d'1 case (selon A.2.) par tour Une partie étant constituée de 20 tours, le joueur aura plus de chance, en tout, d'avancer de 20 cases. Le plateau pourrait être constitué de 20 cases, sachant qu'un joueur peut également avancer de plus d'1 case par tour.
TB17	Je conseillerai à Monsieur Dupond de faire plusieurs simulations sur tableur. Ce qui lui permettra d'avoir une moyenne de cases.

TA18	Le jeu se passe en 20 tours, on avance entre 0 et 3 cases. Si on avance toujours de 3 cases on a avancer de 60 cases au maximum et au minimum de 0 cases. Si on ajoute la somme des cases en fonction du nombre de cases que l'on avance : $3\times20 + 2 \times 20 + 1\times20 + 0 \times 20 = 60+40+20+0 = 120$. Si on fait une moyenne 120 : 2 = 60 cases. Le jeu peu comporter 60 cases.
TA21	Monsieur Dupond pourrait réaliser plusieurs simulations comme celle que nous avons faite précédemment puis faire la moyenne du nombre de cases parcourues en 20 tours.
TA24	Il détermine le nombre de cases du plateau de jeu en fonction de la valeur moyenne de la case A22
TA26	Un joueur chanceux réalisera à chaque tours un écart de 3 or les probabilités de faire un trois est la plus faible. il va donc falloir établir le nombre de cases en fonction de l'écart le plus probable c'est à dire de faire un 1 (écart). Si on se dit qu'en un tour tous les joueurs font 1, il leur faudra 20 tours pour arriver à 30 minutes soit 20 cases.
TA27	Recommencer l'expérience plusieurs fois et faire ensuite la moyenne des résultats obtenus puis choisir cette valeur moyenne comme nombre de cases séparant la case de départ de la case d'arrivée.
TA28	Je conseille Monsieur Dupond de refaire plusieurs simulations avec cette feuille de calcul. Le nombre de case serait la plus petite valeur obtenue dans la cellule (B ;1) [111] pour les simulations.

TA29	Pour conseiller M. Dupont, on notera les résultats obtenus en B21 [112], dont on calculera la moyenne. Cette moyenne sera alors, en arrondissant au nombre entier le plus proche, le nombre de case conseillé pour le jeu de M ; Dupont. (Sinon, pour éviter ces calculs, on pourra « tirer » vers la droite les cases allant de A1B1 à A21B21 jusqu'à Y1Z1 et Y21Z21, par exemple, puis calculer la moyenne des sommes obtenues.
TA30	Pour conseiller Monsieur Dupond, je ferais plusieurs simulations pour obtenir une moyenne du nombre total de cases parcourues par un joueur. Ayant une moyenne faite sur une centaine de simulations (facile

[111] En B1 se trouve pour cet élève la formule qui calcule de combien de cases on a avancé après 20 tours.

[112] En B21 se trouve pour cet élève la formule qui calcule de combien de cases on avancé après 20 tours.

	à obtenir grâce à EXCEL ou OPENOFFICE), Monsieur Dupond saura combien de cases il faut mettre sur le plateau de jeu.
TA31	Je conseille à M.D. de faire la simulation 10 fois, de noter les résultats, d'en effectuer la moyenne. Il aura là l'espérance de la case d'arrivée au bout de 20 tours, pour un joueur.
TA32	Monsieur Dupond pourrait répéter cette simulation un certain nombre de fois en notant à chaque fois le nombre de cases parcourues par un joueur au bout de 20 tours, et calculer ensuite la moyenne de ces résultats, déterminant ainsi le nombre moyen de cases parcourues par un joueur après 20 tours. S'il répète cette expérience un grand nombre de fois, ce nombre moyen sera proche de l'espérance du nombre de cases parcourues par un joueur en 20 tours, c'est à dire en 20 minutes. Monsieur Dupont pourra choisir ce nombre comme nombre de cases de son jeu.
TA33	choisir un nombre de cases proche de 20 pour que le joueur pouvant avancer de 2 cases de plus n'atteigne pas la case d'arrivée trop tôt et pour que le joueur ne pouvant avancer son pion que de zéro ou 1 case (par « manque de chance » ne désespère pas quant à ses chances de gagner la partie.

PB45	X*barre* = $\sum$ *i de 1 à k* P[X=x_i] = $0 \times \frac{4}{10} + \frac{3}{10} + 2 \times \frac{2}{10} = 3 \times \frac{1}{10}$ $= \frac{7}{10}$ On a 20 tours donc il faut que le jeu ne soit pas plus long que $20 \times \frac{7}{10}$ = 14 cases.
PA46	En simulant 3 parties avec 20 tours, et en faisant la moyenne de nombre des cases trouvées, on obtient une valeur comprise entre 20 et 29 cases. Il faudrait que Mr. Dupond simule plusieurs parties afin d'obtenir une moyenne de cases plus précises.
PA47	Il devra prendre un nb de case [20 ; 30]
PA48	Il devrait mettre environ 25 de cases pour être sûr que les joueurs finissent la partie en 30 minutes à 3
PA49	Monsieur Dupond devrai mettre 20 cases sur son jeu.
PA50	Pour conseiller M. Dupond, il faut voir le nombre de case dont on va avancer en suivant les probabilité. On calcule l'espérance du lancer : $0 \times \frac{1}{4} + 1 \times \frac{3}{8} + 2 \times \frac{1}{4} + 3 \times \frac{1}{4} = \frac{3}{8} = \frac{4}{8} + \frac{3}{8} = \frac{10}{8}$ On multiplie ce « résultat moyen » par le nombre de tour : $\frac{10}{8} \times 20 = 25$. Sur un grand nombre de partie, les résultats des joueurs tendent vers 25 cases. Pour que le jeu puisse être souvent gagner, le plateau de jeu doit faire 25 cases.
PA51	Il faut 25 cases.
PA53	Il faut qu'il simule un grand nombre de fois 3 fois 20 lancers de dés et qu'ensuite il fasse la moyenne des sommes des 3 fois 20 lancers. On obtient un intervalle [20, 30] avec un pic autour de 25
PA54	Je conseille à Dupond de faire 4 case.
PA55	Il faudrait qu'il mette environ 32 cases pour finir une partie en 20 minutes.
PA56	Il faudrait que Monsieur Dupond fasse un jeu de 25 cases pour que le jeu dure 20 tours maximum.

TA57	Soit E(X) l'espérance : $E(X) = (0\times1/4)+(1\times3/8)+(2\times1/4)+(3\times1/8) = 5/4 = 1,25 > 0$. L'espérance étant supérieure à 0, il peut espéré donc de gagné dans le temps disposé. Je luis conseille de joué avec bonne conscience.
TA58	Il pourrait prendre l'écart du milieu (2) et multiplier par 20 (car il y a 20 tours). Il pourrait faire un jeu avec 40 cases.
TA59	On lui suggère de calculer l'espérance de gain, ou de répéter les expériences un grand nombre de fois pour obtenir une valeur l'approchant. Puis on détermine de combien de cases a progressé un joueur en 20 tours. On obtient la valeur moyenne pour laquelle un joueur peut finir la partie.
TA60	Pour conseiller Monsieur Dupond il faudrait répéter l'expérience sur tableur un grand nombre de fois, la loi des grands nombres nous dit que la valeur de l'avancée totale devrait se rapprocher de l'espérance de gain. Pour que le jeu finisse en 20 tours, Monsieur Dupond devrait prendre cette valeur estimée en cases soit 25 cases (calcul de l'espérance au dos) Soit E l'espérance de gain en 20 tours $E(X) = (4/16 \times 0 + 6/16 \times 1 + 4/16 \times 2 + 2/16 \times 3) \times 20 = 25$
TA61	On observe la case D20 qui est le nombre total de cases avancée par le joueur. On répète de très nombreuses fois cette simulation, en notant à chaque fois le résultat obtenu en D20. Puis on fera la moyenne de ces résultats. Cela nous donnera donc le nombre moyen de case avancés par un joueur en plusieurs parties.
TA62	Pour que Monsieur Dupond choisisse le nombre de cases de son jeu, il faut qu'il fasse un très grand nombre de simulations (plusieurs milliers) et qu'il calcule la moyenne du total des cases parcourues par un joueur en 20 tours. d'après la loi des grands nombres, cette moyenne est environ égale à l'espérance du nombre de cases parcourues en 20 tours. Monsieur Dupond n'a alors plus qu'à prendre la partie entière de la moyenne qui vient d'être calculée comme nombre de cases de son jeu.
TA63	Il faut relancer beaucoup de fois les calculs en appuyant sur « F9 » ($\approx$ 1000 fois), puis il faut prendre la moyenne des résultats de D1 comme nombre de cases.
TA64	Pour l'aider à prendre sa décision concernant le nombre de cases séparant la case de départ de la case d'arrivée, je conseillerais à Monsieur Dupond de calculer l'espérance du nombre de cases que le joueur peut obtenire en 1 lancer puis de multiplier celle-ci par le nombre de lancers qui est ici 20. Le résultat obtenus pourrait-être pris comme nombre de cases. $E(X) = \sum(X).\ P(X)$
TA65	Je calcule l'espérance $E(X) = 0\times1/4 + 1\times3/8 + 2\times1/4 + 1/8\times3$ $= 5/4$ $= 1,25$ Sur un tour le joueur à l'espérance de parcourir 1,25 cases. Donc sur 20 tours $1,25\times20 = 25$ Le joueur peut parcourir 25 cases. Ainsi Mr Dupond devrait inclure 25 cases pour que son jeu dure une demi-heure.
TA66	M. Dupond devrait relancer un très grand nombre de fois (par ex : 1000 fois) la simulation pour avoir la moyenne du nombre de cases qu'un joueur aura avancé à la fin de la partie pour avoir une idée du

	nombre de cases séparant la case de départ, de la case d'arrivée.
TA67	Monsieur Dupond devrait simuler un grand nombre de fois le jeu, puis faire la moyenne du nombre de cases obtenues. Et enfin prendre la partie entière de cette moyenne, si celle-ci est décimale afin d'avoir une idée du nombre de cases à mettre.
TA68	Monsieur Dupond peut utiliser le simulateur et ~~calculer~~ noter les différents résultats du nombre de cases avancées après 20 tours. La moyenne de ces résultats-ci lui donnerait une quantité de cases de son jeu qui ~~assureraient~~ *illisible* le temps de la partie dans les environs de 30 minutes.

TA69	Je conseillerai à Monsieur Dupond d'ajouter une dizaine de case car ceci permettra d'avoir un peu plus de point et ceci définiera le gagnant.
TA71	Je conseillerais a Monsieur Dupond de mettre un nombre de case supérieur à 20 pour son jeu mais qui ne dépasse pas 25 cases.
TA72	Le tableau montre que le joueur a plus de chance de se déplacer d'une case car sa probabilité est la plus grande. Il y a au maximum 20 tours. Si le jeu est composé de 20 cases alors il y a une forte probabilité que tous les joueurs finissent la partie. il y aura alors certainement un joueur qui finira la partie avant la fin des 20 tours.
TA73	Pour prendre sa décision concernant le nombre de cases séparant la case de départ de la case d'arivée, je conseillerais à Monsieur Dupont de faire varier plusieurs fois les calculs pour trouver la valeur de E2 qui correspond au nombre de déplacements effectués pendant une partie. Cette valeur étant expérimentale, je lui conseillerais d'augmenter un peu le nombre de cases pour être sur qu'il suffira pour n'importe quel jeu.
TB74	Calculons E l'espérance du nombre de cas : $E(X) = 0\times4/16+1\times6/16+2\times4/16+3\times2/16$ $E(X) = 6/16+8/16+6/16$ $E(X) = 20/16$ $E(X) \approx 4/3 \approx 1{,}25$ A chaque tour, un joueur peut espérer avancer de 1,25 cases. Donc pour 20 tours, il peut espérer avancer de 25 cases. Monsieur Dupond devrait donc faire un plateau de 25 cases pour espérer que la partie finisse en 30 minutes.
TB75	M. Dupond devrait mettre un nombre de cases supérieur à 40 cases sachant que le gain 1 (de case) a le plus de probabilité de tomber et que le gain 2 et 3 a une probabilité de 4/16.
TB76	Si le joueur obtient 3 cases à chaque tour, il avancera de $3\times20 = 60$ cases au bout de 20 tours.
TB80	*Les probabilités des écarts 0, 1, 2 et 3 viennent d'être calculées : 4/10, 3/10, 2/10 et 1/10* 8/20 ; 6/20 ; 4/20 ; 2/20 → on sait qu'il y a 20 tours d'où le 20.

<table>
<tr><td></td><td><table><tr><td>Pendant 8 tours, le joueur n'avance pas.
Pendant 6 tours, le joueur avance d'une case
Pendant 4 tours, le oueur avance de deux cases
Pendant 2 tours, le joueur avance de trois cases
Il faudrait donc 20 cases pour ce jeu, soit une case par tour.</td><td>O cases
6 cases
8 cases + 6 cases = 14 cases
6 cases + 14 cases = 20 cases</td></tr></table></td></tr>
<tr><td>TB81</td><td>Monsieur Dupond doit calculer l'espérance du nombre de cases parcourues par le joueur en un tour, et multiplier ce résultat par 20 pour connaître l'espérance du nombre de cases parcourues par le joueur en 20 tours. Le résultat obtenu sera le nombre de cases à intégrer à son jeu de plateau.
Ici E(cases parcourues par tour) = 6/16+2/4+3/8 = 5/4, soit 1,25 cases par tour.
$20 \times 1,25 = 25$
Donc M. Dupond devrait mettre 25 cases sur son jeu pour espérer avoir un gagnant après 20 tours.</td></tr>
<tr><td>TB82</td><td>Pour décider du nombre de cases de ce jeu, il faut calculer l'espérance de cases à chaque tour et puis le faire pour au maximum 20 tours.
$E(X) = 0\times4/16+1\times6/16+2\times4/16+3\times2/16 = \frac{6+8+6}{16} = 5/4$
nbre de cases : $5/4 \times 20 = 25$
M. Dupond devrait concevoir un jeu avec 25 cases pour qu'il y ait 1 gagnant au maximum au bout de 20 tours.</td></tr>
<tr><td>TB83</td><td>Il faut calculer l'espérance de X
$E(X) = 0\times1/4+1\times3/8+2\times1/4+3\times1/8$
$= 3/8+4/8+3/8$
$= 10/8$
$= 1,25$
On peut espérer avancer d'une case par tour, donc M. Dupond devrait faire un plateau composé de 19 cases séparant le départ de l'arrivée pour qu'au 20ème tour le pion arrive sur la case arrivé.
Départ Arrivée↓
dessin d'un pion</td></tr>
<tr><td>TB84</td><td>$E(X)=p(0)\times0+p(1)\times1+p(2)\times2+p(3)\times3$
$= 3/8 + ½ + 3/8$
$= 10/8$
$= 1,25$
Les joueurs peuvent espérer avancer de « 1,25 » case par tour, soit ils peuvent espérer avancer de $1,25 \times 20 = 25$ cases en 20 tours. Ainsi je conseillerais à Monsieur Dupond de mettre sa case d'arrivée 25 cases plus loin de sa case de départ.</td></tr>
<tr><td>TB85</td><td>Pour conseiller Monsieur Dupond, il faut calculer l'espérance :
$E(X) = 2/8\times0 + 3/8\times1 + 2\times2/8 + 3\times1/8 = 1,25$.
L'espérance est positive donc il a plus de chance de gagner mais l'espérance n'est pas très importante alors jouer entraine des risques qu'il perd et qu'il natteigne pas la case d'arrivée. Maintenant c'est à M. Dupont de choisir.</td></tr>
</table>

TB86	E(X) = 6/16 + 8/16 + 6/16 = 20/16 Au bout de 16 tours, le joueur avancera de 20 cases environ puis pour les quatres derniers tours (le joueur avancera d'environ 4 tours.) Le jeu devrait être constitué de 20 (à 25) cases.
TB87	*Le calcul de l'espérance sur 20 tours pour un joueur a été effectué dans la question A.2. qui précédait (il n'était pas demandé dans l'énoncé).* Grâce au calcul de l'espérance, on sait que chaque joueur évolue en moyenne de 1,25 cases. Si l'on multiplie ce nombre par 20 qui représente le nombre de tours possibles pour qu'un joueur gagne dans le délai imparti de 30 minutes, on constate que 1,25 × 20 = 25. Ainsi, le nombre de cases entre la case de départ et celle d'arrivée est proche de 25 cases.
TB88	Il faut calculer l'espérance du nombre de cases que le joueur peut faire en lançant 2 déq et la multiplier par 20 pour avoir le nombre de cases que le joueur peut espérer atteindre en 20 tours. E(X) = 0×1/4+1×3/8+2×1/4+3×1/8 = 5/4 de cases. 20 E(X) = 20 × 5/4 = 25 cases. Le nombre de cases séparant la case de départ de la case d'arrivée doit être de 25 cases.
TB89	*Cette élève a commencé la question A.1. en simulant 3 joueurs : en A1, B1 et C1 des libellés : « Premier joueur », etc, et en A2, B2 et C2 la formule ABS(ALEA.ENTRE.BORNES(1 ;4) – ALEA.ENTRE.BORNES(1 ; 4)). On peut penser que le choix d'une formule condensée pour cette élève était une nécessité qui a pu s'imposer lorsqu'elle a voulu gérer les trois joueurs. Elle a rapidement réalisé qu'elle était hors sujet (c'est dommage, il eut été intéressant de voir la suite…), et nous a demandé une autre feuille pour recommencer.* • On calcule l'espérance pour un tour d'un joueur, cette espérance représente l'espérance du nombre de case qu'il peut parcourir E(X) = 0×2/8+1×3/8+2×2/8+3×1/8 = 3/8+4/8+3/8 = 10/8 = 1,25 (cases) • En multipliant cette espérance valable pour un tour, nous trouverons l'espérance de nombre de cases parcourues à la fin des 20 tours : E(X) × 20 = 1,25 × 20 = 25 (cases) On peut donc considérer que le nombre de cases « idéales » parcourues au bout de ces 30 mn par un joueur est 25 cases. Cependant pour plus de sécurité il serait préférable que Monsieur Dupont choisisse un nombre plus faible de case. En effet nous ne savons pas à quelle amplitude ce nombre de cases varies globalement (écart-type ?) et l'amplitude maximale des valeurs est importante : 3 × 20 – 0 × 20 = 60. Même si la probabilité de n'avoir que des écarts égaux à 0 ou à 3 est plus faible, celle d'avoir une espérance juste un peu inférieur à 25 sur 20 parties n'est pas négligeable et sur les trois joueurs il est tout à fait probable que les 3 ont une espérance inférieure à 25. Donc selon moi il est préférable d'abaisser le nombre de cases à 20 pour que la probabilité que sur les 3 joueurs au moins l'un atteigne le bout du plateau et par conséquent la fin du jeu soit plus « sûre » c'est à dire plus élevée Bien sûr le plateau où il est sûr que les joueurs atteignent la fin : celui à zéro case fait perdre au jeu son intérêt, c'est pour cela qu'il vaut mieux éviter de diminuer trop sensiblement le nombre de cases.

TB90	*L'élève a calculé dans la question A.2. l'espérance pour un tour, qui n'était pas demandée.* Monsieur Dupond pourrait calculer le nombre de cases avancée en 20 tours avec l'espérance trouvée dans le A.2. Il revient à faire : $1{,}25 \times 20 = 25$, on en déduit qu'il peut choisir un parcours de 25 cases pour terminer le jeu en 30 minutes. Pour être sur de ne pas déborder des 30 minutes, il peut choisir moins de 25 cases pour son parcours. Il peut aussi, pour trouver une moyenne de cases avancée *illisible* tours, lancer la feuille de calcul du A.1. un grand nombre de fois (> 1000) et ainsi faire la moyenne de toutes les sommes obtenues.
PA92	La valeur affichée en C21 est le nombre de cases qu'a parcourues le joueur en 20 tours. Je conseille ainsi à M. Dupond de réitérer plusieurs fois ce calcul (avec d'autres aléatoires), et de faire la moyenne des différentes valeurs qu'il aura trouvées. Avec 10 tests comme celui-ci effectués, M. Dupond aurait une valeur approchée du nombre moyen de cases parcourues par un joueur après 20 tours. Pour obtenir une certaine assurance que le jeu se termine après au maximum 20 tours, il peut ôter à cette moyenne quelques cases (par exemple entre 5 et 10), et ce nombre final conviendrait bien à son problème car il aura de grandes chances se termine avant 20 tours.
PA94	Selon moi, il faudrait qu'entre la case de départ et d'arrivée, il y ait aurait 20 cases au moins pour qu'il y ait au moins un gagnant.
PA96	Moyenne de X : $E(X) = \frac{1}{4} \times 0 + \frac{3}{8} \times 1 + \frac{1}{4} \times 2 + \frac{1}{8} \times 3 = 1{,}25$ En moyenne, un joueur se déplacera de 1,25 cases par tour. Il y a 20 tours donc il faudra 1,25 x 20 = 25 cases sur le plateau pour qu'un joueur puisse espérer finir. Pour rajouter de la difficulté, on peut augmenter le nombre de cases jusqu'à 60 (si un joueur avance de 3 à chaque tour) mais finir le jeu sera très difficile.
PA97	$E(X) = 0 \times \frac{1}{4} + 1 \times \frac{3}{8} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} = \frac{3+4+3}{8} = \frac{10}{8} = \frac{5}{4}$. L'espérance est de $\frac{5}{4}$ $E(X) \times 20 = \frac{100}{4} = 25$. Le nombre de cases moyen parcourues par un joueur en 20 tours est donc de 25. nombre de tours Monsieur Dupond devrait donc créer son jeu avec 25 cases, case d'arrivée incluse.
PA98	Monsieur Dupond devrait faire un grand nombre de simulations (par exemple 100) et calculer la médiane de cette série * car la médiane serait plus appropriée que la moyenne (celle-ci étant influencée par les extrêmes) * la série du total de cases parcourues.
PA99	Il fait un tableur comme précédemment et fait varier les lancés (plusieurs fois). Il retient les sommes inscrites dans la cellule D2 et peut ainsi voir (à peu près) le nombre de cases dont il doit munir son jeu.
PA101	Monsieur Dupont devrait réaliser plusieurs fois l'expérience avec le tableur un grand nombre de fois et faire une moyenne du nombre totale de cases parcourues et en arrondissant cette moyenne pour avoir le nombre de case de son jeu. cela d'après la théorie des grands nombres.
PA103	Valeur de X x_0 x_1 x_2 x_3 0 1 2 3

	$P(X=x_i)$ $\quad \frac{2}{8} \quad \frac{3}{8} \quad \frac{2}{8} \quad \frac{1}{8}$ $E(X) = \frac{2\times0 + 1\times3 + 2\times2 + 3\times1}{8} = \frac{10}{8} = \frac{5}{4}$ Un joueur peut espérer avancer $\frac{5}{4}$ de case par tour. $\frac{5}{4} \times 20 = \frac{5\times4\times5}{4} = 25$. Il faudrait conseiller à M. Dupond de prévoir 25 cases pour que le premier joueur gagne en 20 tours et que la partie dure 30 min.
PA104	Il doit trouver quelle somme apparaît le plus souvent dans la case C21, qui correspond à la somme des cases avancées par le joueur. Monsieur Dupond devrait simuler une dizaine de parties de 20 tours et relever à chaque fois la somme affichée en C21. Ensuite il calcule la moyenne de ces nombres, et l'arrondit par défaut à l'unité près. Si on appelle S la somme en C21, il a calculé l'espérance de cette variable.
PA105	$Total_X$ = Nombre total de cases = Nb tours × X or X varie sur I = [0 ; 4] on s'intéresse à $x = \frac{3+4+3}{40} = \frac{1}{4}$ $Total_X = 20 \times x = 20 \times \frac{1}{4} = 5$ Il faudrait 7 cases pour éviter que la partie ne finisse par trop rapidement si jamais le joueur obtient 3 de façon récurrente.
PB106	Sur 20 tour un joueur obtient environ 8 fois le nombre 1 ($\frac{3}{8} \times 20 = 7{,}5$) ; 5 fois le nombre 0 ; 5 fois le nombre 2 et 2 fois le nombre 3 ($20 \times \frac{1}{8} = 2{,}5$). Soit un joueur avance d'environ 8 + 5×2 + 3×2 = 22 cases. Le plateau doit contenir environs 22 cases.
PB109	Nous savons qu'un tour permet d'avancer en moyenne de $\frac{7}{4}$ cases → $3\times\frac{4}{16} + 2\times\frac{6}{16} + \frac{4}{16} = \frac{28}{16} = \frac{7}{4}$ En moyenne, un joueur avancera donc de 35 cases. $\frac{7}{4} \times 20 = \frac{140}{4} = 35$. Pour être certain qu'il y aura un gagnant au bout de 20 tours (ou 30 min), il faudra donc que M. Dupond choisiss un nombre de cases inférieur à 35.
PB110	A chaque tour il peut avancer de 1, 2 ou 3 cases dans le cas où il avancerait de 3 cases à chaque tour il avancerait au bout des 20 tours de 60 cases Mais la probabilité la plus importante est d'avancer d'aucune case le nombre de case : $\frac{3\times1 + 3\times2 + 3\times1}{1+2+3} \times 20 = 33{,}3$ donc 34 cases pour gagner
PB111	$E(X) = 0 \times \frac{1}{4} + 1 \times \frac{3}{8} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8}$ $= \frac{3}{8} + \frac{4}{8} + \frac{3}{8} = \frac{10}{8} = \frac{5}{4}$

PB114	Si nous considérons toujours 20 tours et que à chaque tour l'écart est de 3 le résultat est de 60. L'idéal serait de prendre 45 cases (soit le 3ème quartile) pour que le jeu ne dure ni trop longtemps ni trop peu de temps
PB115	Au meilleur des cas, un joueur fait toujours X = 3. En 20 tours, il fait donc 20 × 3 = 60 cases. Cependant la probabilité qu'en ces 20 tours, il parcourt 60 cases est quasi-nulle. Au moins bon cas, le joueur fera 20 × 0 = 0 cases. Puusque 1 est le score le plus probable, le jeu doit contenir plus de 20 cases. Puisque cette probabilité est la plus haute, le joueur gagnera en fonction des autres scores.
PA116	$E(X) = x_1 \times p_1 + x_2 \times p_2 + x_3 \times p_3$ $= 0 \times \frac{2}{8} + 1 \times \frac{2}{8} + 2 \times \frac{2}{8} + 3 \times \frac{2}{8}$ $= 1{,}5$
PB118	$E = 0 \times \frac{1}{4} + \frac{3}{8} \times 1 + 2 \times \frac{1}{4} + \frac{1}{8} \times 3 = = \frac{3+3+4}{8} = \frac{10}{8} = \frac{5}{4} = 1{,}25$ Comme l'espérance est de 1,25 le joueur a plus de chance de faire A au dé. La case de départ ne compte pas comme une case : c'est le point 0 Pour atteindre la case d'arrivée, qui compte comme une case, en 20 tours le jeu doit comporter 21 cases où la case 21 est l'arrivée et la 1 le départ. Car le joueur risque d'avancer d'1 case par tour donc de 20 cases rn 20 tours.
PB119	$E(s) = \frac{1}{4} \times 0 + \frac{3}{8} \times 1 + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} = 1{,}25$ → En moyenne un joueur avance donc de 1,25 case par tour. 1,25 × 20 = 25 En 20 lancés de dé un joueur avance en moyenne de 25 cases. ⇒ M. Dupond devrait donc faire un plateau d'au moins 30 voire 35 cases car ce résultat n'est qu'une moyenne, il se peut donc qu'un joueur chanceux gagne en moins de 20 tours et donc en moins de 30 min.
PB120	Le joueur avance en moyenne d'une case à chaque tour comme il en fait 20 le nombre de cases doit être de 20 environ. mais comme il est possible qu'un des joueurs n'ai « pas de chance » et que le nombre maximal de tours est de 20, il est plus prudent que le nombre de cases soit inférieur à 20 pour finir la partie dans le temps imparti.
PB121	En 30 minutes le joueur peut avancer au maximum de 60 cases et au minimum de 0 case, je pense qu'il devrait faire une moyenne enre ces deux données.
PB122	en calculant l'espérance E de la variable aléatoire X.
PB123	Une partie de mr Dupond dure au maximum 20 tours. Soit E l'espérance d'avancer de ce jeu. $E = \frac{1}{4} \times 0 + \frac{3}{8} \times 1 + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} = \frac{1}{2} + \frac{6}{8} = \frac{5}{4} = 1{,}25$. 1,25 × 20 = 25 Il doit donc y avoir au maximum 25 cases. Mais dans ce cas, le jeu pourrait se terminer au 20^{e} tour. Avec 20 cases (donc 18 entre la case de départ et celle d'arrivée) le jeu sera toujours terminé avant le dernier

	tour.
PB124	Grâce à l'espérance E(X) on peut trouver le nombre moyen de cases pour lesquelles le joueur avance $E(X) = \sum x_i\, p_i \sim 1{,}2$
PB125	Je pense qu'il devrait y avoir 60 cases au maximum, car une personne très chanceuse qui aura toujours un écart de 3 entre les valeurs des dés qu'il a lancé, jouera 20 fois (come tous les autres donc 3 × 20 = 60 ; il y aura 60 case (maximum) Espérance de gain :
PB126	Ayant 30 cas favorables, je lui suggèrerais 30 cases séparant la case de départ de la case d'arrivée.
PB127	<u>Calculons l'espérance E(X)</u> : $E(X) = \sum p_i\, x_i = 0 \times \frac{4}{14} + 1\times \frac{6}{16} + 2 \times \frac{4}{16} + 3 \times \frac{2}{16} = (\ldots) = \frac{5}{4}$ Donc au bout d'un tour, un joueur éspère avancer d'un peu plus d'une case. Au bout de 20 tours, il devrait espérer avancer de $\frac{5}{4} \times 20 = 25$ cases. Pour que Monsieur Dupond puisse espérer qu'il y ait un gagnant au bout de 20 tours, il serait astucieux de placer 25 cases entre la case de départ et la case d'arrivée.
PB128	• Dans un 1er temps, je calculerais l'espérance mathématique. Soit $E(X) = \sum p_i\, x_i$ $= (\ldots)$ $= \frac{5}{4}$ • Puis après je calculerais je multiplierais cette varience mathématique par 20 (soit 20 tours) Ce qui nous donne E(x) × 20. Pour conseiller Monsieur Dupond *illisible : mauvais scan*
PB130	Calculons l'espérance de déplacement du joueur lors d'un tour : $E = 0 \times p(0) + 1 \times p(1) + 2 \times p(2) + 3 \times p(3)$ $= 0 \times 4/16 + 1\times 6/16 + 2 \times 4/16 + 3 \times 2/16 = 20/16 = 1{,}5$ *faute de calcul* Monsieur Dupond estime qu'en 30 minute s'effectue 20 tours. Le joueur peut donc espérer avancer d'un nombre de case égal à 20 E, c'est à dire 20 × 1,5 = 30 cases. On peut donc conseiller à M. Dupond de placer 30 cases entre l'arrivé et le départ s'il veut que le jeu dure 30 minutes. On remarque que l'écart type esr de $\sqrt{5} \sim 2{,}23$ $(\sigma = \sqrt{(p(0) - E)^2 + (p(1) - E)^2 + (p(2) - E)^2 + (p(3) - E)^2}\ )$ *formule fausse* Ce qui est un peu élevé, on peut donc conseiller à M. Dupond de mettre un peu moins de 30 cases s'il veut que le jeu dur 30 minutes ou moins de manière catégorique.
PB131	espérance : $E(X) = \sum (x_i\, p_i) = 1{,}25$ Le nombre de case séparant celle de départ et cele d'arrivée varie selon le nombre de joueur donc on ne peut pas conseiller Monsieur Dupond puisqu'on peut jouer à 1 comme à 3 joueurs (cela n'est pas possible)
PB133	Selon la loi des grands nombres, sur 20 tours un joueur ne devrait pas avancer à 8 reprise, avancer de 1 case à 6 reprise, avancer de 2 cases à 4 reprises, avancer de 3 cases à 1 reprise. Donc en 30 minutes, un joueur devrait avancer de 20 cases. La partie ne devrait pas durer plus de 30 minutes, et le jeu se jouant à 2 ou 3 joueurs, Mr Dupond devrait

	mettre 7 cases à son jeu pour espérer avoir un gagnant dans les délais impartits.
PB134	Je calculerais l'espérance € de la loi de probabilité précédente. on a E = 0 × 4/16 + 1 × 6/16 + 2 × 4/16 + 3 × 2/16 = 1,25. Si l'on multiplie cette espérance par 20 on obtient un jeu avec 25 cases séparant le départ de l'arrivée. On remarque de plus une probabilité importante d'obtenir un 1. Si l'on multiplie par 20 tours on obtient un jeu avec 20 cases séparant le départ de l'arrivée. Je lui conseillerias donc de faire un plateau entre 20 et 25 cases.
PB135	Etant donné que l'on connaît la probabilité pour chacune des issues on peut déterminer le nombre de case que devrait environ parcourir un joueur pour 20 tours. On va donc multiplier le numérateur de la probabilité de X auquel il correspond et faire la somme de ces résultat pour tous les X, Cet a dire : 8×0 + 1×6 + 2×4 + 3×2 = 28 Pour une partie de 20 tours les joueurs devront avancer de plus ou moins 28 cases. C'est donc le nombre que l'on va conseiller à monsieur Dupond. Car de temps à autre des joueurs feront plus de 28 cases ou moins de 28 . *La loi de probabilité de cet élève faisait apparaître à l'origine des fractions au dénominateur 20, ce qui explique peut-être ce raisonnement.*
PB136	Il est plus probable d'obtenir un écart entre les points amenés égal à 1 (avancer d'une case) que d'obtenir un écart entre les points amenés égal à 5. Autrement dit, il serait judicieux de réduire le nombre de cases séparant la case départ de la case arrivée.
PB137	On calcule l'espérance de la série statistique. E(X) = 0 × 4/16 + 1 × 6/16 + 2 × 4/16 + 3×2/16 = (…) = 1,25 Ainsi le joueur qui lance les dés peut espérer avoir au moins 1 et même plus. On divise le nombre de tours par l'espérance 20/1,25 = 16. Ainsi si Monsieur Dupond met 16 cases il pourra espérer avoir un gagnant en vingt tours.
PB138	Calcul de l'espérance : E(X) = (…) = 1,25 A chaque tour, le joueur espère avancer de 1,25 cases, soit 5 cases en 4 tours. Donc pour déterminer le nombre de cases, il doit effectuer le calcul : $\text{tps un tour d'1 joueur} = \frac{1{,}5 \text{ min}}{3} = 0{,}5 \text{ min}$ $\text{nombre cases} = \frac{\text{tps total}}{\text{n tps d'un tour d'1 joueur}} \times E(X)$ $= \frac{30 \text{ min}}{n \times 0{,}5 \text{ min}} \times 1{,}25$ avec n ∈ [1 ; 3] correspondant au nombre de joueur.
PB139	$E(x) = \sum x_i\, p_i$ = (…) = 1,25 pour un tour l'espérance est d'avancer de 1,25 cases. Ainsi si l'on fait 20 tours, l'espérance est de 25 cases. Donc pour conseiller monsieur Dupont, il faut ~~au moins~~ 25 cases environ entre la case de départ et la case d'arrivée pour qu'il y ait un gagnant dans les délais impartis.
PB140	On calcule l'espérance grâce au résultat précédent : $E = \sum x_i\, p_i$ = (…) = 1,25.

	On estime, en probabilité, le gain moyen de cases par tour par joueur proche de 1,25. En théorie, sachant qu'il y a 20 tours au plus, il faudrait que le nombre de cases idéal séparant la première de la dernière case du plateau soit 25 cases (20 × E) ou (20 × 1,25). Cela à partir de la case de départ, c'est à dire qu'il faudrait 26 cases au total, avec la case de départ numérotée 1.
PA141	Monsieur Dupond répète de nombreuses fois l'opération (1000 fois) il fait la moyenne des sommes obtenues. Cette moyenne correspondra au nombre de cases possibles pour espérer qu'il y ai un gagnant. Sur les trois joueurs, un atteindra avant les autres cette moyenne au bout de 20 tours ou légèrement plus. Cette moyenne correspondra à l'espérance mathématique de X. $E(X) = 0 \times 4/16 + 6/16 + 2\times4/16 + 3\times2/16 = 20/16 = 5/4 \sim 1$. On peut donc dire que Mr Dupond peut donner à son jeu 20 cases de la case de départ à celle d'arrivée.
PA142	$E(X) = 0 \times 2/8 + 1 \times 3/8 + 2 \times 2/8 + 3 \times 1/8$ $= 3/8 + 4/8 + 3/8$ $= 10/8$ $= 5/4$ L'espérance de cette loi est 5/4. Il y a 20 tours donc : $5/4 \times 20 = 100/4 = 25$. Je conseillerais à Mr Dupond de faire son jeu avec 25 cases s'il veut espérer avoir un gagnant.
PA143	Il pourrait calculer l'espérance de X pour savoir le nombre de cases que peut espérer un joueur avancé en 1 tour puis multiplier ce résultat par 20. Cela lui donnera le nombre de cases que peut espérer un joueur avancer en une partie de 20 tours. Il utilisera alors ce nombre comme nombre de cases du jeu.
PA144	Il faudrait qu'il réalise plusieurs fois la simulation fait dans le tableur ci-dessus, afin de faire une moyenne de la somme, qui est le nombre de case mené lors des 20 tours, pour décider du nombre de cases qui séparent la case de départ et la case d'arrivée. Ainsi en faisant la moyenne de ces échantillons, il permettra au jeu de ne pas être ni trop facil, ni trop dur d'atteindre avec le temps réparti. Puis on divise le nombre de case par le nombre de joueur puisque le tableur ci-dessus n'est que le temps repartie pour un joueur.
PA145	Supposons qu'un des joueurs obtienne 3 comme écrat entre les points amenés à chaque tour ; et si au bout de 20 tours il se trouve à la case d'arrivée. Il faut donc $3 \times 20 = 60$ cases pour qu'il y arrive ; Monsieur Dupond doit donc prendre 59 cases ou moins, pour qu'un joueur puisse espérer gagner et pour séparer la case de départ et la case d'arrivée.
PA146	En utilisant la somme, M[r] Dupond aura un premier ordre d'idée quand au nombre de cases nécessaire et suffisant. Ensuite grâce à la moyenne et aux quartiles il verra quelles valeurs sont le plus souvent prises, le nombre de cases sera évalué en fonction de cela. Avec ces informations, il pourra évaluer le nombre de cases pour que le jeu se termine en moins de 20 tours, mais pas de manière trop rapide également.
PA147	Il faudrait qu'il y ait 20 cases, car la probabilité de faire avancer son pion d'une case est plus importante que les autres (6/16). Or, pour finir dans les délais impartis (20 tours), il faudrait $1 \times 20 = 20$ cases.
PA148	On calcul l'espérance de X, tel que X représente le nombre moyen de cases parcourus pour chacun des joueurs en l'espace de 10 lancés donc de 10 tours : $E(X) = \sum x_i p_i$

	$E(X) = 0 \times 4/10 + 1 \times 3/10 + 2 \times 2/10 + 3/10$ $E(X) = 10/10 = 1$ Soit en 10 tours en moyenne chaque joueur avancerais de 1 case, je conseillerai donc a M. Dupond de limiter le nombre de cases à 2 si il ne veut pas dépasser la limite de 30 minutes. Cependant je dirai aussi a M. Dupond que plus de de tours est important, plus il y a de chance que ces calculs soit proche de la réalité. Etant donné que le nombre de 20 tours au plus est imposé je lui dirai que il est probable d'avoir dans la réalité des résultats différents des miens.
PA149	Poue aider M. Dupond à prendre sa décision quant au nombre de cases de son nouveau jeu, on calcule l'espérance pour un tour de jeu, c'est à dire le nombre de cases que peut espérer avancer un joueur : $E(X) = \sum p_i x_i = 1{,}25$ cases. On multiplie ensuite cette espérance par 20 (qui est le nombre de tours par personne dans le jeu) Ce qui nous donne : $1{,}25 \times 20 = 25$ cases. Un joueur peut espérer avancer 25 cases en 20 tours. Donc M. Dupond aurait intérêt à choisir comme nombre de cases 25 pour son jeu pour une durée de 20 tours en 30 minutes.
PA150	Monsieur Dupond peut utiliser la formule MAX dans son tableur pour savoir le nombre de point maximum qu'un joueur peut obtenir. Le nombre de cases de son jeux sera donc inférieur ou égale à ce nombre maximum.
PA152	Je lui conseillerai de calculer le nombre de cases à mettre en fonction du temps imparti et de la probabilité d'avancer d'un certain nombre de cases.
PA153	$E(X) = \sum x_i p_i = 0 \times 4/16 + 1 \times 6/16 + 2 \times 4/16 + 3 \times 3/16 = 1{,}25$ $1{,}25 \times 20 = 25$ A ce jeu, en 20 tours, un joueur peut espérer avancer de 25 cases. Il faudrait donc 24 cases entre la case de départ et la case d'arrivée pour qu'un joueur puisse espérer gagner.
PA154	Je conseillerai à M. Dupond de faire 20 cases car la plus grande probabilité est d'avancer d'une case âr tour et qu'il souhaite faire durer la partie au plus 20 tours, au minimum les joueurs auront donc parcourus 20 cases. Il est aussi possible de répéter un très grand nombre de fois la feuille de calcul du tableur (A.2), d'observer la somme obtenue en D1 pour constater que pour chaque expérience le joueur aura au minimum avancer de 20 cases.
PA155	Pour prendre sa décision concernant le nombre de cases séparant la case de départ de la case d'arrivée, Monsieur Dupond pourrait faire une multitude, beaucoup de simulations de parties de 20 tours en ne s'intéressant qu'au nombre de cases parcourues au bout de ces 20 tours. Monsieur Dupond pourrait ensuite établir une moyene de tous les résultats obtenus. Cette moyene serait donc le nombre de cases que compterait le jeu : de la case de départ à la case d'arrivée.
PA156	Pour que monsieur Dupont choisisse sa décision concernant le nombrede cases du jeux, il faudrait déjà que celui-ci calcule l'espérance de gain d'un joueur durant 1 tour. On appelle cette espérance de gain E(x). et $E(x) = 0 \times 1/16 + 6/16 + 2 \times 4/16 + 3 \times 2/16 = 20/16 = 1{,}25$ points par tour en moyenne. Donc le joueur va en moyenne gagner A,25 points en 1,5 minute.

	Ce qui signifie qu'en 30 mn, il va gagner1,25 × 30/15 = 25 points durant une partie en moyenne et sachant qu'à chaque tour il gagne 1,25 points, en une partie, le nombre de cases doit être de 25/1,25 = 20 cases.
PA157	Je conseille à Monsieur dupont de faire son jeu sur 20.

****

Annexes du chapitre 21

Annexe 13 : projets de programmes pour la classe de terminale S (rentrée 2012)

3. Probabilités et statistiques

On approfondit le travail en probabilités et statistiques mené les années précédentes.
Afin de traiter les champs de problèmes associés aux données continues, on introduit les lois de probabilités à densité. Le programme en propose quelques exemples et, en particulier, la loi normale qui permet notamment d'initier les élèves à la statistique inférentielle par la détermination d'un intervalle de confiance d'une proportion à un niveau de confiance de 95 %.
Cette partie se prête particulièrement à l'étude de problèmes issus d'autres disciplines.
Le recours aux représentations graphiques et aux simulations est indispensable.

CONTENUS	CAPACITÉS ATTENDUES	COMMENTAIRES
Conditionnement, indépendance Conditionnement par un événement de probabilité non nulle. Notation $P_A(B)$.	• Construire un arbre pondéré en lien avec une situation donnée. • Exploiter la lecture d'un arbre pondéré pour déterminer des probabilités. • Calculer la probabilité d'un événement connaissant ses probabilités conditionnelles relatives à une partition de l'univers.	On représente une situation à l'aide d'un arbre pondéré ou d'un tableau. On énonce et on justifie les règles de construction et d'utilisation des arbres pondérés. Un arbre pondéré correctement construit constitue une preuve. Le vocabulaire lié à la formule des probabilités totales n'est pas un attendu du programme, mais la mise en œuvre de cette formule doit être maîtrisée.
Indépendance de deux événements.	▣ Démontrer que si deux événements A et B sont indépendants, alors il en est de même pour $\overline{A}$ et B.	Cette partie du programme se prête particulièrement à l'étude de situations concrètes. ◇ Des activités algorithmiques sont menées dans ce cadre, notamment pour simuler une marche aléatoire. ⇆ [SVT] Hérédité, génétique, risque génétique.
Notion de loi à densité à partir d'exemples Loi à densité sur un intervalle.		Les exemples étudiés s'appuient sur une expérience aléatoire et un univers associé Ω, muni d'une probabilité. On définit alors une variable aléatoire X, fonction de Ω dans $\mathbf{R}$, qui associe à chaque issue un nombre réel d'un intervalle I de $\mathbf{R}$. On admet que X satisfait aux conditions qui permettent de définir la probabilité de l'événement $\{X \in J\}$ comme aire du domaine : $\{M(x,y)\,;\, x \in J \text{ et } 0 \leq y \leq f(x)\}$ où f désigne la fonction de densité de la loi et J un intervalle inclus dans I. Toute théorie générale des lois à densité et des intégrales sur un intervalle non borné est exclue.

Ministère de l'éducation nationale, de la jeunesse et de la vie associative – Direction générale de l'enseignement scolaire
Projet de programme de mathématiques en classe terminale de la série S
Page 12 sur 17

CONTENUS	CAPACITÉS ATTENDUES	COMMENTAIRES
Loi uniforme sur $[a, b]$. Espérance d'une variable aléatoire suivant une loi uniforme.	• Connaître la fonction de densité de la loi uniforme sur $[a, b]$.	L'instruction « nombre aléatoire » d'un logiciel ou d'une calculatrice permet d'introduire la loi uniforme sur $[0,1]$. La notion d'espérance d'une variable aléatoire à densité sur $[a, b]$ est introduite à cette occasion par $\int_a^b t\,f(t)\,\mathrm{d}t$. On note que cette définition constitue un prolongement dans le cadre continu de l'espérance d'une variable aléatoire discrète. AP *Méthode de Monte-Carlo.*
Loi exponentielle.	• Calculer une probabilité dans le cadre de la loi exponentielle.	▣ On montre qu'une variable aléatoire T suivant une loi exponentielle vérifie la propriété de durée de vie sans vieillissement : pour tous réels t et h positifs, $P_{T \geq t}(T \geq t+h) = P(T \geq h)$.
Espérance d'une variable aléatoire suivant une loi exponentielle.	▣ Démontrer que l'espérance d'une variable aléatoire suivant une loi exponentielle de paramètre λ est $\frac{1}{\lambda}$.	L'espérance est définie comme la limite quand x tend vers $+\infty$ de $\int_0^x t\,f(t)\,\mathrm{d}t$ où f est la fonction de densité de la loi exponentielle. Cette partie du programme se prête particulièrement à l'étude de situations concrètes, par exemple sur la radioactivité ou la durée de fonctionnement d'un système non soumis à un phénomène d'usure.
Loi normale centrée réduite $\mathcal{N}(0,1)$. Théorème de Moivre Laplace (admis).	• Connaître la fonction de densité de la loi normale $\mathcal{N}(0,1)$ et sa représentation graphique. ▣ Démontrer que pour $\alpha \in \left]0,1\right[$, il existe un unique réel positif t_α tel que $P(-t_\alpha \leq X \leq t_\alpha) = 1-\alpha$ lorsque X suit la loi normale $\mathcal{N}(0,1)$. • Connaître les valeurs approchées $t_{0,05} \approx 1{,}96$ et $t_{0,01} \approx 2{,}58$.	Pour introduire la loi normale $\mathcal{N}(0,1)$, on s'appuie sur l'observation des représentations graphiques de la loi de la variable aléatoire $Z_n = \frac{X_n - np}{\sqrt{np(1-p)}}$ où X_n suit la loi binomiale $\mathcal{B}(n, p)$, et cela pour de grandes valeurs de n et une valeur de p fixée entre 0 et 1. Le théorème de Moivre Laplace assure que pour tous réels a et b, $P(Z_n \in [a,b])$ tend vers $\int_a^b \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x$ lorsque n tend vers $+\infty$.

CONTENUS	CAPACITÉS ATTENDUES	COMMENTAIRES
Loi normale $\mathcal{N}(\mu, \sigma)$ d'espérance μ et d'écart-type σ.	• Utiliser une calculatrice ou un tableur pour calculer une probabilité dans le cadre d'une loi normale $\mathcal{N}(\mu, \sigma)$.	Une variable aléatoire X suit une loi $\mathcal{N}(\mu, \sigma)$ si $\dfrac{X-\mu}{\sigma}$ suit la loi normale $\mathcal{N}(0,1)$. On fait percevoir l'information apportée par la valeur de l'écart-type. ⇆ [SI et SPC] Mesures physiques sur un système réel en essai.
       Estimation Intervalle de fluctuation.	• Connaître une valeur approchée de la probabilité des événements suivants : $\{X \in [\mu-\sigma, \mu+\sigma]\}$, $\{X \in [\mu-2\sigma, \mu+2\sigma]\}$ et $\{X \in [\mu-3\sigma, \mu+3\sigma]\}$, lorsque X suit la loi normale $\mathcal{N}(\mu, \sigma)$. ◙ Démontrer que si la variable aléatoire X_n suit la loi $\mathcal{B}(n, p)$, alors, pour tout α dans $]0,1[$ on a, $\lim\limits_{n\to+\infty} P\left(\dfrac{X_n}{n} \in I_n\right) = 1-\alpha$, où I_n désigne l'intervalle $\left[p - t_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}, p + t_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$. • Connaître l'intervalle de fluctuation asymptotique(*) au seuil de 95 % : $\left[p - 1{,}96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}, p + 1{,}96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ où p désigne la proportion dans la population.	La connaissance d'une expression algébrique de la fonction de densité de la loi $\mathcal{N}(\mu, \sigma)$ n'est pas un attendu du programme. On illustre ces nouvelles notions par des exemples issus des autres disciplines. La démonstration ci-contre donne l'expression d'un intervalle de fluctuation asymptotique(*) au seuil $1-\alpha$ de la variable aléatoire $F_n = \dfrac{X_n}{n}$, qui à tout échantillon de taille n associe la fréquence obtenue. En pratique, on fait l'approximation dès que $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$. En majorant $1{,}96\sqrt{p(1-p)}$, on retrouve l'intervalle de fluctuation présenté en classe de seconde. La problématique de prise de décision, déjà rencontrée, est travaillée à nouveau.

CONTENUS	CAPACITÉS ATTENDUES	COMMENTAIRES
Intervalle de confiance (*). Niveau de confiance.	• Estimer une proportion inconnue à partir d'un échantillon. • Déterminer une taille d'échantillon suffisante pour obtenir une estimation d'une proportion au niveau de confiance 0,95 pour une précision attendue.	Les attendus de ce paragraphe sont modestes et sont à exploiter en lien avec les autres disciplines. ◙ On peut démontrer que l'intervalle $\left[F_n - \frac{1}{\sqrt{n}}, F_n + \frac{1}{\sqrt{n}}\right]$ contient la proportion à estimer avec une probabilité au moins égale à 0,95. On peut alors énoncer que p est élément de l'intervalle $\left[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}\right]$ avec un niveau de confiance de plus de 95 %, où f désigne la fréquence observée sur un échantillon de taille n. La simulation de sondages sur tableur permet de sensibiliser aux fourchettes de sondage. Il est important de noter que, dans d'autres champs, on utilise l'intervalle $\left[f - 1{,}96\frac{\sqrt{f(1-f)}}{\sqrt{n}}, f + 1{,}96\frac{\sqrt{f(1-f)}}{\sqrt{n}}\right]$ qu'il n'est pas possible de justifier dans ce programme. ⇆ [SVT] Analyse de graphiques où les données sont fournies par des intervalles de confiance. (AP) *Prise de décision lors de la comparaison de deux proportions (par exemple lors d'un essai thérapeutique).*

(*)Avec les notations définies précédemment :
Un intervalle de fluctuation asymptotique de la variable aléatoire F_n au seuil $1-\alpha$ est un intervalle déterminé à partir de p et de n et qui contient F_n avec une probabilité d'autant plus proche de $1-\alpha$ que n est grand.

Un intervalle de confiance pour une proportion p à un niveau de confiance $1-\alpha$ est la réalisation, à partir d'un échantillon, d'un intervalle aléatoire contenant la proportion p avec une probabilité supérieure ou égale à $1-\alpha$, intervalle aléatoire déterminé à partir de la variable aléatoire F_n, qui à tout échantillon de taille n associe la fréquence.
Les intervalles de confiance considérés ici sont centrés en f.